ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 8, 58–83 (1972)

Three Sources of Criterion Variance: Static Dimensionality, Dynamic Dimensionality, and Individual Dimensionality[1]

ANDRES INN AND CHARLES L. HULIN
University of Illinois, Urbana

AND

LEDYARD TUCKER
University of North Carolina, Chapel Hill

This paper examines traditional usage of criterion measures and finds these traditional methods inadequate for retaining the lawful variance of the measures. Three sources of lawful variance are investigated, and evidence is found for the multidimensionality of each source: static dimensionality, dynamic dimensionality, and individual dimensionality. The dimensional complexity of the criteria coupled with traditional usage which discards lawful variance and the use of inappropriate predictors may contribute to the discouraging validities reported in the prediction of performance. The paper outlines a procedure which analyzes the sources of criterion variance and proposes two possible methods for criterion usage which may retain more of the lawful variance.

Organizational decisions require a criterion. Raises, promotions, firings, transfers, etc., are usually based on a criterion of effective performance. What is needed, then, is one criterion score which represents the extent of effective performance for each person within the organization. Each person's score could then be compared to a cutting score; if his score exceeds the cutting score he is promoted, or he receives a raise in salary, or some other positive consequence. If, however, his score is below the cutting score, he is fired, demoted, or receives negative consequences from the organization. Multiple criteria could be used but this only complicates the problem by a factor related to the number of criteria used.

An obvious and frequently used solution involves rating scales. Management can simply rate each employee on the extent to which he meets the

[1] This article is condensed from a thesis by the senior author submitted to the Department of Psychology at the University of Illinois in partial fulfillment of the requirements for the M.A. degree.

criterion of effective performance. Even if one adequately solves the difficult problem of defining effective performance in a meaningful way (Smith & Kendall, 1963), rating scales are still subject to criticism. Response sets, chance response tendencies and rater biases may influence the scores. These criticisms are epecially grave in light of Title 7 of the 1964 Civil Rights Act which expressly forbids any employer

> to limit, segregate, or classify his employees in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

Clearly, a criterion of performance which is easily subject to such biases might be illegal.

The obvious direction to proceed involves obtaining an objective measure of job performance. But a quick glance through the literature should convince one that there is certainly no consensus as to which measure is most inappropriate; there are almost as many measures of job performance as there are investigators utilizing such measures. The numerous measures possible and the lack of consensus outline a definition problem. That is, what is the best way to define job performance? Does one simply pick one measure and utilize it alone, or does he take a number of measures on the criterion?

Measurement theory suggests that one cannot presume an isomorphism between a measure and the criterion. Different measures have different components of true, method, and error variance. With a reliance on a single measure the degree of correspondence between the measure and the inferred trait or characteristic is unknown. On the other hand, by using additional measures and a triangulation process, increasing evidence on the correspondence between the measure and the inferred trait or construct is gained. At the same time, additional information as to the error component is acquired. This argument for convergent validity has been aptly expressed by Campbell and Fiske (1959). An integral part of validation by convergence and triangulation is the requirement for some evidence of relative validity demonstrated by greater variance shared within a given trait than is shared by the methods of measurement. The most appropriate solution to criterion measurement appears to be the use of multiple measures with orthogonal error variances.

This is an important decision. By using a number of criterion measures one constructs a nomological network and essentially defines the construct, job performance, as the entire network of interrelationships among the measures. But, again, it is important to remember that the measures will be used to make decisions and there are a number of alternative ways in which the measures may be used: (1) An investigator may select one of his measures on the basis of its relationships with other measures and deal solely with it, (2) he may choose to combine his criterion measures into one single composite score, (3) he may choose to relate each predictor variable to all of the criterion variables, or (4) he may choose to apply a factor analytic technique and deal with variance common to a number of measures.

Single Criterion Measure

The first alternative is not a viable choice. The investigator who chooses the criterion measure on the basis of its relationship with a predictor would be completely ignoring the error component which might be solely responsible for the predictor-criterion relationship. Lack of knowledge about the error component violates all of the reasoning behind taking multiple measurements in the first place.

Another shortcoming is evident with the use of only one performance measure. Many different behaviors define job performance and it is difficult to conceive of a single measure which adequately samples all of the relevant behaviors. In other words, the measure would probably not be central to the nomological net defining the construct of job performance. Thus, a single measure can be, at best, deficient, or at worst, completely irrelevant if it has a large error component or taps only behaviors unnecessary for the job.

Uncertainty as to the behavior of the error component renders a single performance measure useless for portraying change over time. It is certain that learning and increased experience on the job will function to change performance over time. Other kinds of change are also to be expected over time; pilots exhibit radically different behaviors in the time interval between take-off and touch-down, and college professors function differently during holidays and exam periods than during midsemester. Seasonal changes affect the jobs of farmers, ranchers, contractors, lifeguards, ski instructors, and weathermen. A single performance measure may confirm that a change in performance occurred, but it is never certain whether it is the error component which changes or the true score component.

Likewise, a single performance measure will tend to mask individual differences in performance. People can be expected to react differently to the same job situation since they have unique combinations of abilities and experience. Otis (1940) has suggested that different employees make different contributions within an organization. Consider a sales clerk whose salesmanship is so overpowering that her performance is consistently profitable. Another clerk might build good will through her politeness and courteous attention, thus encouraging customers to continue their purchases throughout the store. It is just as feasible that the error component of a single measure would behave differently for different people as it would be for the true score component to behave differently for different people. For instance, some persons might be subject to more measurement error than others. Discrimination among people with respect to differences in performance on a single measure would be difficult as one could never be certain that the discrimination among persons was a function of real differences or error.

Single Composites

There are a number of ways in which criterion measures may be combined so as to form a single composite criterion. Any composite criterion is usually the sum of the weighted variables and competitive methods utilize unique configurations of weight. The choice of weights is necessarily confounded with the psychological meaning of the criterion and therefore the generality of the criterion. Another frequently used method for combining criterion measures involves assigning unit weights and summing the measures. This method, however, equates the measures on importance which may not be desirable. For example, one might find it difficult to assume that revenue earned and number of times late are equally important.

Of course, one may choose any configuration of weights for the measures so that he may derive a score Y_i for each ith person by multiplying his score on the jth measure by its appropriate weight and summing over n measures:

$$Y_i = \sum_{j=1}^{n} w_j \cdot x_{ij}.$$

To arrive at a useful criterion score, Y, it is essential that one may discriminate among persons on the basis of the composite score. That is, one may choose weights for the measures so as to maximize the variance of Y. Horst (1936) presents such a method for calculating weights for measures so as to maximize differences between all possible pairs of subjects, thereby maximizing the variance of the derived composite.

An equally attractive solution is presented by Edgerton and Kolbe (1936). These investigators assume that each of the criterion measures is a measure of the same construct, job performance, and therefore for each individual the difference between weighted measures (in standard score form) should be small. That is,

$$d_{ij} = w_i x_i - w_j x_j,$$

and weights are chosen so as to minimize d_{ij} for all possible combinations of i and j.

Ghiselli (1956) states that both the Edgerton and Kolbe and Horst methods yield identical results in that both weight individual criterion measures in terms of their loadings on the first principal component. This again might not be an appropriate solution. For instance, if the first principal component accounts for 42% of the variance and the second component accounts for 38% of the variance, is it appropriate to completely discount the second component?

The combination of measures into a single composite can be approached from still another angle. If each of the criterion variables could be expressed in terms of a common dimension then the combination would involve merely the computation of an unweighted sum. This simple operation is suggested by Brogden and Taylor (1950) in applying cost accounting procedures to the concept of criterion construction. These investigators advocate the conversion of production units, errors, time, etc., into dollar units. While this initially appears to be a tidy solution to an extremely complex problem the difficulties of applying cost accounting procedures to all performance aspects of all jobs are immense. Consider the difficulties of assessing the dollar value to the job performance of a systems analyst or to the various job activities of an executive. Any attempt to relate the more intangible aspects of performance to a dollar criterion would require as complex a solution as the selection of weights.

The methods for constructing single composite criteria mentioned above do demonstrate one advantage over the use of a single criterion measure; they require multiple measurement. By carefully selecting his measures an investigator can employ construct validation procedures to gain more information on the correspondence between his measures and the construct. In addition, multiple measures allow for more effective sampling of the relevant behaviors on the job, and the investigator can be more confident that his measures are tapping the pertinent behavior patterns or performance dimensions. However, the single composite methods lose information by combining measures. Even though each measure may reflect a different pattern of job behavior or performance dimension, the composite score reduces these to a single dimension. The consequence is to mask both change in performance over time and the effects of individual differences on the job. Consider a job which has both a quality and quantity dimension of employee productivity with equally weighted measures which correspond to these dimensions. If in his initial months on the job an employee concentrates on quality rather than quantity and receives scores of 5 and 1, respectively on these measures, the composite score would be 6. But, if he were to reverse his emphasis from quality to quantity and the respective measures were now 1 and 5, the composite score would indicate no change. Individuals who view the job differently and adopt either a quality or quantity oriented behavior pattern could receive identical scores on the composite criterion. Discrimination between these individuals would be impossible and prediction of variables which are so global as to have little behavioral meaning may prove equally impossible.

These methods of composite criterion construction assume that each measure obtained by the investigator represents an imperfect outcropping of the same underlying trait. Otherwise, while mathematically and statistically correct, the procedure makes no sense psychologically. Combining measures with zero covariances which measure independent traits will result in a composite measure of limited or no psychometric value. The assumption should be a matter for empirical study.

Retaining the Multiple Criteria

Dunnette (1963), in his discussion of the criterion problem, expressed dissatisfaction with attempts to combine multiple criteria. He advocates the third approach—an investigation of the relationships between predictors and the various criteria.

Dunnette claims that validity is lost in the process of combining multiple criteria to form a single performance measure. If each of the measures correlate highly, the investigator can be more assured that his measures are really reflecting the same construct and weighting to achieve a single measure of the construct would be appropriate. However, the higher the correlations between measures, the less the necessity to weight the individual measures to form a single composite. The argument works conversely. The lower the correlations between independent measures, the more the investigator can be sure that he is measuring independent dimensions and the rationale for combining qualitatively different measures becomes weaker. A composite derived from independent measures is not representative of any particular dimension and produces what Dunnette calls a "distilled essence" measure. Dunnette suggests investigating and summarizing the relationships among several predictors and several criteria instead of trying to isolate one variable or dimension and calling it performance.

Little information is lost by summarizing relationships among predictors and criteria. The multiple measures allow both construct validity and effective sampling of criterion dimensions. Behavior changes over time as well as the effects of individual differences can be traced from the separate criterion measures. While the correlations among predictors and criteria lose little information, they are also relatively useless for summarizing and organizing information as well as for decision making. One must not lose sight of the fact that criteria and an analysis of patterns of interrelationships between predictors and criteria can be of little value in making decisions. Dunnette's critique of weighting solutions appears closely related to a defense of clinical judgments over actuarial judgments, a position which in light of several investigations (Wiggins & Hoffman, 1968) seems untenable.

Factor Analytic Approaches

It appears to be inappropriate to use a single measure, a combined criterion score, or the measures separately. Instead, one may wish to investigate the dimensionality of the construct, job performance, and utilize the structure of the concept to derive his criterion scores. There are, however, three possible multidimensional sources of variance defined by the construct of job performance: static dimensionality, dynamic dimensionality, and individual dimensionality (Ghiselli, 1956).

Factor analysis is an empirical technique which may be used to investigate the dimensionality of performance measures due to a number of different modes or data sources. For example, Grant (1955), Seashore and Tiffin, and Rush (reported in Lawler, 1967) have factored objective performance measures taken at one point in time. Such studies indicate that from 3 to 5 factors may be expected to account reasonably for the variance in performance measures taken at one point in time. However, to the extent that performance can be expected to change over time, and to the extent that individuals load differentially on these factors, criteria based on the static structure of criterion measures may not be adequate.

Factor analysis may also be harnessed to investigate variance of a criterion measure due to time. That is, instead of factoring a matrix of variables from one time period, one factors a matrix of time periods to discover the dimensions which "explain" variance of measures due to time. Again, to the extent that the different time periods form independent dimensions and to the extent that individuals are differently affected by change due to time, criteria based on the static structure of criterion measures may be inadequate.

Variance due to individual differences or individual dimensionality may be investigated similarly. In this case, one factors a matrix of individuals to discover the dimensions which represent characteristic patterns of behavior. And, as might be expected, to the extent one may expect these characteristic behavior patterns to change over time and over the different performance measures, criteria based on the individual differences structure of criterion measures may be inadequate.

Obviously, it would be desirable to investigate and summarize the dimensionality of each source of criterion variance simultaneously. This paper presents such an investigation. The main purpose of this investigation is to illustrate the usefulness of a multivariate technique for studying the dimensionality of a set of measures. No claim will be made regarding the generality of the substantive findings. On the contrary, the data to be presented are quite likely unique in that they represent a large number of *objective* performance measures assessed at a number of consecutive points in time on a large number of employees performing *identical* tasks. Thus, the data represent a nearly ideal situation for performance assessment. To the extent that these data yield results which indicate complex dimensionality then similar results might well be obtained if the dimensionality of other sets of objective performance assessments are studied.

METHOD

The criterion measures for this analysis were 11 performance measures collected from 184 reservations agents in the Midwest office of a large domestic airline. The reservations agents perform identical functions; each agent acting as an intermediary between the customer on the telephone and the computerized reservations system. The employees receive a month of intensive training emphasizing conversations, answers to queries, and essential questions needed to effect a sale. In addition, as each agent has a keyboard directly linked to the central computer, he is necessarily familiar with the computer format needed to properly record a transaction. Calls are automatically distributed to agents so that the person who has been waiting longest receives the next call.

The performance measures for each agent included: (1) the average amount of dollar revenue earned per 8-hr shift, (2) the average number of tickets sold through travel agents, airport ticket agents, teletype, or other airlines per shift,[2] (3) the average number of telephone calls answered per shift, (4) the average number of tickets sold by will-call arrangement per shift, (5) the average number of tickets sold by mail per shift, (6) the average number of sales leads generated per shift, (7) the average amount of dollar revenue earned by following up previous sales leads per shift, (8) the supervisor's rating of "Customer Impact," (9) the percentage of accurate entries into the airline's computer, (10) the average number of times late per shift, and (11) the average number of hours absent per shift. Each of the 11 measures were collected during 5 consecutive

[2] Agents are instructed to encourage customers to purchase tickets before their actual departure. Tickets are available in advance either by mail, by will-call arrangement, or through travel agents, airport ticket agents, teletype, or other airlines. The three measures dealing with ticketing reflect the manner in which tickets actually were purchased by an agent's customers.

months—November and December of 1967, and January, February, and March of 1968.

With the exception of the supervisor rating, the performance data are objective. Revenue and ticketing figures are tallied automatically by the computer reservations system and supplied to management in a monthly printout. Automatic counters keep track of incoming telephone calls made to each agent and the agents record these figures themselves. The number of times late and number of hours absent are computed from time cards. Sales leads are records of potential customers; these are recorded on forms and are tallied by supervisors. The computer accuracy score is derived from random sampling of agents' input and indicates the percentage of correct entries. These performance data are exceptional compared to the usual criterion measures available. The 11 performance measures were selected so as to have no linear dependencies, and correlated sampling error may be discounted as only two of the measures (supervisor rating and the computer accuracy score) are a result of sampling. All other measures reflect the sales efforts of each employee for the duration of the month.

Some of these data were not available for each employee for each of the five months due to rotating vacations, turnover, and periodic hiring. A missing data generation suggested by Ledyard Tucker was applied to the data in order to complete the data matrix. The technique is based on successive factoring of the cross-product matrix and the simultaneous estimation of scores based on the factor structure within each iteration.[3]

The original data are characterized by meaningful mean differences among the variables while variance differences were not meaningful. To remove variance differences the $184 \times 11 \times 5$ matrix was collapsed over the 5 time periods to yield a matrix 920×11. The standard deviation of each performance variable measured for the 184 agents over the 5 time periods was computed and each entry in the 920×11 matrix was divided by its standard deviation. Of the 11 performance variables only the Customer Impact rating did not have a meaningful zero point and consequently this variable was standardized over the 920 observations.

Any score in the final, scaled, three-dimensional ($184 \times 11 \times 5$) data matrix can be cross-classified by each of the three dimensions or *modes.* That is, the scores can be classified in terms of scores for the 184 subjects, scores on each of the 11 variables, and scores during each of the 5 months. Furthermore, Tucker's (1966) three-mode factor analysis model states

[3] Anyone interested in a more detailed description of the method or the FORTRAN program used to generate the missing data points may address the senior author at the Department of Psychology, University of Illinois, Champaign, ILL 61820.

that it is possible to approximate the data matrix from factor matrices corresponding to the different modes:

$$X_{ijk} = [{}_jC_p \times {}_kT_q]_{(p\cdot q)}G_mS_i, \tag{1}$$

where C is a matrix of criterion variable factors, T is a matrix of time factors, S is a matrix of subject factors, and G is the core matrix which interrelates C, T, and S. The subscripts i, j, and k refer to the number of observations in each mode. In the present study $i = 184$, $j = 11$, $k = 5$. The m, p, and q refer to the number of dimensions used to account for variance in each mode. The accuracy of the approximation and the psychological interpretations attached to the data depend on the number and nature of the factors retained to explain variance in each mode.

The suitability of the model to the analysis requirements stated above is readily apparent. Static dimensionality (criterion variable factors), dynamic dimensionality (time factors), and individual dmensionality (subject factors) are each represented as are the interrelationships among the three sources of criterion variance. The sample size of 184 persons is adequate for the three-mode analysis because it is possible to collapse over modes. That is, static dimensionality may be investigated by factoring the sum of squares and cross products (SSCP) matrix of 920 observations (184 subjects within each of 5 time periods) of the 11 criterion measures. Dynamic dimensionality may be investigated by factoring the cross product matrix of 2024 observations (11 measures for each of 184 subjects) of the 5 time periods, and individual dimensionality may be investigated by factoring the cross product matrix of 184 observations of 55 variables (11 measures within each of the 5 time periods).

RESULTS

Static Dimensionality

The C matrix of Tucker's model is formed by factoring the SSCP matrix of 920 observations of 11 variables. Table 1 presents the roots and percent of variance accounted for with each successive factoring of the matrix.[4] Three performance dimensions seem to account adequately for the variance in the criterion measures taken at one point in time. This three-dimensional space approximates 98.5% of the sum of squares of

[4] The large proportion of variance accounted for by the first dimension is a function of factoring a SSCP matrix rather than a correlation matrix. Correlations are based on standard scores which have eliminated mean and variance differences among measures and have therefore discarded approximately 90% of the meaningful variance among these measures. The first factor of a SSCP matrix extracts variance due to mean differences and therefore accounts for most of the variance discarded from a correlation matrix.

TABLE 1
VARIANCE ACCOUNTED FOR BY AN n-DIMENSIONAL APPROXIMATION OF PERFORMANCE MEASURE VARIANCE

n-Dimensional space	Roots	Percentage of variance	Cumulative percentage
1	260376	96.7	96.7
2	3489	1.3	98.0
3	1271	.5	98.5
4	989	.4	98.8
5	797	.3	99.1
6	681	.3	99.4
7	578	.2	99.6
8	470	.2	99.8
9	407	.2	99.9
10	194	.1	100.0
11	22	.0	100.0

the 920×11 array and the sum of squares of discrepancies accounts for the remaining 1.5%. Table 2 presents the rotated eigenvectors for the three dimensional solution.[5]

It is evident from the rotated solution in Table 2 that variables which

TABLE 2
VARIMAX ROTATED EIGENVECTORS FOR THE 3-DIMENSIONAL APPROXIMATION OF CRITERION DIMENSIONALITY

	Vectors		
Variables	Speed and accuracy	Customer rapport	Sales ability
1 Dollar revenue	.04	−.05	.49
2 Tickets sold	.07	−.07	.42
3 Calls answered	.36	.04	.04
4 Will-call tickets sold	.04	.16	.32
5 Mailed tickets	−.00	−.05	.46
6 No. of sales leads	.04	.68	.01
7 Revenue from sales leads	.04	.66	−.04
8 Customer impact rating	−.01	.20	.01
9 Computer accuracy score	.92	−.07	−.05
10 Times late	−.04	−.01	.41
11 Hours absent	−.07	.14	.28

[5] Anyone interested in the inter-vector plots for criterion measures, time, or subjects may address a request to the senior author at the Department of Psychology, University of Illinois, Champaign, ILL 61820.

TABLE 3
VARIANCE ACCOUNTED FOR BY A n-DIMENSIONAL APPROXIMATION OF TIME VARIANCE

n-Dimensional space	Roots	Percentage of variance	Cumulative percentage
1	263045	97.7	97.7
2	2240	.8	98.5
3	1809	.7	99.2
4	1152	.4	99.6
5	1030	.4	100.0

tap an agent's *Speed and Accuracy* occur together on the first dimension, variables measuring *Customer Rapport* occur together on the second dimension and measures reflecting *Sales Ability* load highly on the third dimension.

Dynamic Dimensionality

Variance due to time was investigated by factoring the SSCP matrix of 2024 observations of scores during the 5 months. Table 3 presents the roots and percent of variance accounted for with each successive factoring of the matrix. A three-dimensional approximation was chosen to account for the meaningful variance due to time. The three dimensional approximation accounts for 99% of the sum of squares of the 5×5 SSCP matrix, and the sum of squares of discrepancies accounts for the remaining 1%. Table 4 presents the unrotated eigenvectors for the three-dimensional solution.

An examination of the plots among the unrotated time dimensions suggests a rotation to ease interpretation of the three dimensions. It was judged desirable to retain the first dimension in its original form so as to

TABLE 4
UNROTATED EIGENVECTORS FOR THE 3-DIMENSIONAL APPROXIMATION OF TIME VARIANCE

	Vectors		
Month	I	II	III
1 November	.44	.48	.28
2 December	.45	.36	.25
3 January	.44	.12	−.18
4 February	.44	−.16	−.80
5 March	.46	−.78	.43

TABLE 5
TRANSFORMED TIME VECTORS

Month	T_A	T_B	T_C
1 November	.44	.00	−.04
2 December	.45	.08	.04
3 January	.44	.33	.03
4 February	.44	.66	−.03
5 March	.46	.66	1.00

characterize a steady level of performance over time. The second dimension was rotated to show a performance increase over time, and the third dimension was rotated to demonstrate little performance variance during the first four months but to exhibit a gain during the last month (March). The time vectors were rotated solely to ease interpretation. In order to preserve the interrelationships among the factor matrices, the inverse of the transformation matrix which accomplished the rotation was applied to the core matrix G. The original data matrix may be approximated equally well with either the unrotated or rotated time dimensions but the rotated solution allows easier interpretation. Figure 1 summarizes the characteristics of the three rotated time dimensions presented in Table 5.

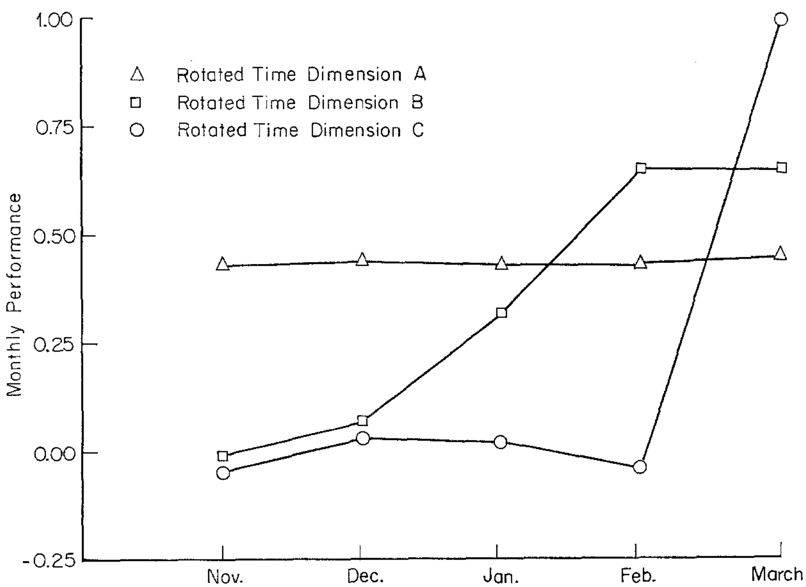

FIG. 1. Monthly performance characterized by the three rotated time dimensions.

Individual Dimensionality

A four-dimensional approximation was chosen to account for individual differences (S matrix) represented by the SSCP matrix of 184 observations of 55 measures. The four-dimensional approximation accounts for 98% of the sum of squares of the original data and the sum of squares of discrepancies accounts for the remaining 2%. Table 6 presents the roots and percent of variance accounted for with each successive factoring of the matrix.

The characteristics of the four subject dimensions may be investigated with reference to the core matrix presented in Table 7. The core matrix is partitioned to present each of the performance dimensions, Speed and Accuracy, Customer Rapport, and Sales Ability with the three rotated time dimensions and the four subject dimensions. Change in performance over time for different kinds of persons may be summarized by Tucker's technique of conceptualizing "idealized individuals" with hypothetical factor loadings. By choosing "idealized individuals" so that their factor loadings are zero on all but the subject dimensions of interest, one may use the performance of the idealized persons to characterize the performance of all persons who load highly on that subject dimension. The characteristic performance of the idealized persons may be determined by performing the matrix operations presented in Tucker's model equation

TABLE 6
VARIANCE ACCOUNTED FOR BY A n-DIMENSIONAL APPROXIMATION OF INDIVIDUAL DIFFERENCES VARIANCE

n-Dimensional space	Roots	Percentage of variance	Cumulative percentage
1	260167	96.6	96.6
2	1749	.6	97.3
3	998	.4	97.6
4	885	.3	98.0
5	641	.2	98.2
6	550	.2	98.4
7	508	.2	98.6
8	443	.1	98.8
9	368	.1	98.9
10	304	.1	99.0
11	290	.1	99.1
12	258	.1	99.2
13	198	.1	99.3
14	178	.1	99.4
.	.	.	.
55	1	.0	100.0

TABLE 7
ROTATED CORE MATRIX

		S_1	S_2	S_3	S_4
Speed and accuracy	T_A	505.74	−7.13	−1.30	−.85
	T_B	−2.08	4.91	−.62	−1.01
	T_C	−1.27	−4.10	−.27	.19
Customer rapport	T_A	19.64	7.11	8.02	1.40
	T_B	16.21	.59	6.54	1.41
	T_C	.54	3.98	−5.47	1.18
Sales ability	T_A	65.14	11.67	.77	38.99
	T_B	3.43	7.36	24.98	32.94
	T_C	3.15	23.88	29.37	−.66

(1). This procedure is illustrated for each of the four subject dimensions.

The first subject dimension is a general factor on which all persons load positively. In order to approximate performance of persons loading on the first subject dimension, one can construct a matrix $\hat{S}_1$ of hypothetical factor loadings for an idealized person a who loads highly on only the first subject dimension:

$$\begin{array}{ccl} & \text{idealized person } a & \\ \hat{S}_1 = & \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{array}{l} \text{loading on } S_1 \\ \text{loading on } S_2 \\ \text{loading on } S_3 \\ \text{loading on } S_4 \end{array} \end{array}$$

Then, the matrix resulting from postmultiplying each of the three partitions of the core matrix G (Table 7), by $\hat{S}_1$ and premultiplying the product by the three time vectors in Table 5, illustrates the performance of persons loading highly on the first subject dimension during the 5 months. Table 8 presents the calculations. The entries in the product matrices presented in Table 8 may be considered indications of the level of performance within each performance dimension during each of the 5 months for idealized person a.

From Table 8 it is evident that a person's loading on the first subject factor corresponds to very little change on performance measures reflecting Speed and Accuracy. That is, the entries in the product matrix representing performance of a person loading only on the first factor demonstrate little change over the 5-month period. It is obvious that different loadings on the first subject dimension will contribute to different levels of performance. Therefore, one may conclude that the level of perform-

TABLE 8

CHARACTERISTIC PERFORMANCE ASSOCIATED WITH THE FIRST SUBJECT DIMENSION

	T_A	T_B	T_C		S_1	S_2	S_3	S_4		Idealized person a	Performance by month of idealized person a
Speed and accuracy											
Nov.	.44	.00	−.04	T_A	505.74	−7.13	−1.30	−.85	S_1	1	222.58
Dec.	.45	.08	.04	T_B	−2.08	4.91	−.62	−1.01	S_2	0	227.37
Jan.	.44	.33	.03	T_C	−1.27	−4.10	−.27	.19	S_3	0	221.80
Feb.	.44	.66	−.03						S_4	0	221.19
Mar.	.46	.66	1.00								230.00
Customer rapport											
Nov.	.44	.00	−.04	T_A	19.64	7.11	8.02	1.40	S_1	1	8.62
Dec.	.45	.08	.04	T_B	16.21	.59	6.54	1.41	S_2	0	10.16
Jan.	.44	.33	.03	T_C	.54	3.98	−5.47	1.18	S_3	0	14.01
Feb.	.44	.66	−.03						S_4	0	19.32
Mar.	.46	.66	1.00								20.27
Sales ability											
Nov.	.44	.00	−.04	T_A	65.14	11.67	.77	38.99	S_1	1	28.54
Dec.	.45	.08	.04	T_B	3.43	7.36	24.98	32.94	S_2	0	29.71
Jan.	.44	.33	.03	T_C	3.15	23.88	29.37	−.66	S_3	0	29.89
Feb.	.44	.66	−.03						S_4	0	35.38
Mar.	.46	.66	1.00								35.38

ance for each subject on the Speed and Accuracy measures is a function of individual differences on the first subject dimension and that the subject's level of performance on these measures remains fairly invariant over time. Likewise, the rank order of subjects on Speed and Accuracy measures is a function of differences in loadings on the first subject dimension, and the rank order among subjects remains stable over time.

Customer Rapport measures, on the other hand, do not exhibit such a degree of stability. Table 8 suggests that subjects on the first subject dimension remain in a stable rank order on measures tapping Customer Rapport, but that the overall level of performance increases at a relatively steady rate. It is also clear that differences in loadings on the first subject dimension contribute to different levels of performance within each month, but for *any* loading on the first subject dimension, the level of Customer Rapport performance increases across time periods from November through March.

Performance on measures reflecting Sales Ability follows a similar pattern of accretion, although the increase in average level of performance during the 5 months is not as great as in the Customer Rapport dimension. Again, individual differences reflected in different loadings on the first subject dimension function to change persons' average levels of performance. The rank order of subjects on the Sales Ability dimension, like the Speed and Accuracy dimension and the Customer Rapport dimension, is not affected by loadings on the first subject dimension.

It is necessary to construct factor loadings for three hypothetical persons in order to characterize the second subject dimension because the second subject factor is bipolar. Idealized persons b, c, and d were located so as to have positive loadings on the first subject dimension but loadings which vary from extremely negative, through zero, to extremely positive on the second subject dimension:

$$\begin{array}{cc}
 & \text{idealized persons} \\
 & \begin{array}{ccc} b & c & d \end{array} \\
\hat{S}_2 = & \left[\begin{array}{ccc} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]
\end{array}
\begin{array}{l} \\ \\ \text{loadings on } S_1 \\ \text{loadings on } S_2 \\ \text{loadings on } S_3 \\ \text{loadings on } S_4. \end{array}$$

The remaining loadings for idealized persons b, c, and d are zero. Again, the matrices resulting from post multiplying each of the three partitioned matrices of G by the new $\hat{S}_2$ and premultiplying the product by the three time vectors will characterize the performance of persons with a pattern of loadings similar to idealized persons b, c, and d. Table 9 presents the calculations.

From Table 9 it is evident that extreme changes in loading on the second subject dimension negligibly affect performance on Speed and Accuracy measures. That is, idealized persons b, c, and d show quite similar patterns of performance throughout the 5-month period on the Speed and Accuracy dimension despite their marked differences in loading. Therefore it is safe to conclude that the second subject dimension has practically no effect on the rank order or the overall level of performance of subjects on Speed and Accuracy measures.

The effects of the second subject dimension on Customer Rapport measures is more marked. Negative loadings, as characterized by idealized person b, function to depress the average level of performance and reverse the general tendency for Customer Rapport performance to improve during the month of March. Positive loadings, on the other hand, tend to raise the general level of performance and preserve the tendency of performance to improve on the Customer Rapport dimension over time.

Performance on the Sales Ability dimension is more affected by the second subject dimension. Negative loadings as typified by idealized person b are associated with a general decrease in level of performance coupled with a decrease in performance over time. Positive loadings as typified by idealized person d are associated with a general increase in level of performance and preserve the tendency of performance to improve over time.

The third subject factor is also bipolar and may be investigated in a similar manner. Three hypothetical persons are used to characterize the third subject dimension:

$$\begin{array}{cc} & \text{idealized persons} \\ & \begin{array}{ccc} e & f & g \end{array} \\ \hat{S}_3 = & \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{array}{l} \text{loadings on } S_1 \\ \text{loadings on } S_2 \\ \text{loadings on } S_3 \\ \text{loadings on } S_4 \end{array} \end{array}$$

The first idealized person, e, loads positively on the first subject dimension and negatively on the third subject dimension. Idealized person f loads positively on only the first subject dimension while idealized person g loads positively on the first and third dimensions. Table 10 presents the matrix manipulations used to characterize the third subject dimension.

Variation in loading on the third subject dimension makes no appreciable difference in performance on the Speed and Accuracy dimension.

TABLE 9

CHARACTERISTIC PERFORMANCE ASSOCIATED WITH THE SECOND SUBJECT DIMENSION

										Idealized persons				Performance by month of idealized persons		
	T_A	T_B	T_C		S_1	S_2	S_3	S_4		b	c	d		b	c	d
Speed and Accuracy																
Nov.	.44	.00	−.04	T_A	505.74	−7.13	−1.30	−.85	S_1	1	1	1		225.55	222.58	219.60
Dec.	.45	.08	.04	T_B	−2.08	4.91	−.62	−1.01	S_2	−1	0	1		230.34	227.37	224.39
Jan.	.44	.33	.03	T_C	−1.27	−4.10	−.27	.19	S_3	0	0	0	=	223.44	221.80	220.16
Feb.	.44	.66	−.03						S_4	0	0	0		220.96	221.19	221.42
Mar.	.46	.66	1.00											234.14	230.00	225.86
Customer Rapport																
Nov.	.44	.00	−.04	T_A	19.64	7.11	8.02	1.40	S_1	1	1	1		5.65	8.62	11.59
Dec.	.45	.08	.04	T_B	16.21	.59	6.54	1.41	S_2	−1	0	1		6.75	10.16	13.56
Jan.	.44	.33	.03	T_C	.54	3.98	−5.47	1.18	S_3	0	0	0	=	10.56	14.01	17.45
Feb.	.44	.66	−.03						S_4	0	0	0		15.93	19.32	22.72
Mar.	.46	.66	1.00											12.63	20.27	27.91
Sales Ability																
Nov.	.44	.00	−.04	T_A	65.14	11.67	.77	38.99	S_1	1	1	1		24.36	28.54	32.72
Dec.	.45	.08	.04	T_B	3.43	7.36	24.98	32.94	S_2	−1	0	1		22.92	29.71	36.51
Jan.	.44	.33	.03	T_C	3.15	23.88	29.37	−.66	S_3	0	0	0	=	21.61	29.89	38.17
Feb.	.44	.66	−.03						S_4	0	0	0		21.55	30.83	40.11
Mar.	.46	.66	1.00											1.27	35.38	69.48

An inspection of Table 10 shows that the performance of idealized persons e, f, and g remains strikingly parallel throughout the 5-month time period. Customer Rapport performance, however, is affected by loadings on the third subject dimension. As one's loading increases from negative through zero to positive, the average level of performance on the Customer Rapport dimension increases. The relationship of loadings on the third subject dimension and Sales Ability is more complex. During November, the greater one's loading on the third subject dimension, the poorer his performance on Sales Ability measures. This relationship reverses during the next four months when an increased loading corresponds to better sales ability performance. In March, this relationship is especially marked and persons with negative loadings on the third subject dimension demonstrate inferior Sales Ability performance while persons loading positively show superior Sales Ability.

The same method was used to investigate the characteristics of the fourth subject dimension. Loadings for three idealized persons, h, i, and j were constructed so as to have each person load positively on the first subject factor but vary from negative, through zero, to positive on the fourth subject factor:

$$\begin{array}{c} \text{idealized persons} \\ \begin{array}{ccc} h & i & j \end{array} \end{array}$$

$$\hat{S}_4 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{loadings on } S_1 \\ \text{loadings on } S_2 \\ \text{loadings on } S_3 \\ \text{loadings on } S_4 \end{array}$$

Table 11 presents the results of the matrix manipulations which characterize the performance of idealized persons h, i, and j.

From Table 11 it is evident that the fourth subject dimension has little effect on Speed and Accuracy performance and only slightly affects Customer Rapport performance. The Speed and Accuracy profiles of persons h, i, and j are quite similar despite the disparity in loadings, while the Customer Rapport performance profiles show only slight improvement in performance as loadings increase from negative to positive. Performance on the Sales Ability dimension, however, is affected substantially by loadings on the fourth subject dimension. Negative loadings, characterized by idealized person h, tend to depress the average level of performance on Sales Ability and reverse the general tendency for performance to improve over the 5-month period. Conversely, positive loadings, characterized by idealized person j, are associated with a somewhat improved level of performance and still preserve the tendency for performance to improve over time.

TABLE 10

CHARACTERISTIC PERFORMANCE ASSOCIATED WITH THE THIRD SUBJECT DIMENSION

										Idealized persons				Performance by month of idealized persons		
	T_A	T_B	T_C		S_1	S_2	S_3	S_4		e	f	g		e	f	g
Speed and Accuracy																
Nov.	.44	.00	−.04	T_A	505.74	−7.13	−1.30	−.85	S_1	1	1	1		223.14	222.58	222.02
Dec.	.45	.08	.04	T_B	−2.08	4.91	−.62	−1.01	S_2	0	0	0		228.01	227.37	226.72
Jan.	.44	.33	.03	T_C	−1.27	−4.10	−.27	.19	S_3	−1	0	1	=	222.59	221.80	221.02
Feb.	.44	.66	−.03						S_4	0	0	0		222.16	221.19	220.22
Mar.	.46	.66	1.00											231.27	230.00	228.72
Customer Rapport																
Nov.	.44	.00	−.04	T_A	19.64	7.11	8.02	1.40	S_1	1	1	1		4.87	8.62	12.37
Dec.	.45	.08	.04	T_B	16.21	.59	6.54	1.41	S_2	0	0	0		6.24	10.16	14.07
Jan.	.44	.33	.03	T_C	.54	3.98	−5.47	1.18	S_3	−1	0	1	=	8.48	14.01	19.53
Feb.	.44	.66	−.03						S_4	0	0	0		11.31	19.32	27.33
Mar.	.46	.66	1.00											17.74	20.27	22.81
Sales Ability																
Nov.	.44	.00	−.04	T_A	65.14	11.67	.77	38.99	S_1	1	1	1		29.59	28.54	27.48
Dec.	.45	.08	.04	T_B	3.43	7.36	24.98	32.94	S_2	0	0	0		26.42	29.71	33.01
Jan.	.44	.33	.03	T_C	3.15	23.88	29.37	−.66	S_3	−1	0	1	=	20.64	29.89	39.13
Feb.	.44	.66	−.03						S_4	0	0	0		15.11	30.83	46.56
Mar.	.46	.66	1.00											−10.60	35.38	81.36

TABLE 11
CHARACTERISTIC PERFORMANCE ASSOCIATED WITH THE FOURTH SUBJECT DIMENSION

										Idealized persons				Performance by month of idealized persons		
	T_A	T_B	T_C		S_1	S_2	S_3	S_4		h	i	j		h	i	j
Speed and Accuracy																
Nov.	.44	.00	−.04	T_A	505.74	−7.13	−1.30	−.85	S_1	1	1	1		222.96	222.58	222.19
Dec.	.45	.08	.04	T_B	−2.08	4.91	−.62	−1.01	S_2	0	0	0		227.82	227.37	221.91
Jan.	.44	.33	.03	T_C	−1.27	−4.10	−.27	.19	S_3	0	0	0	=	222.50	221.80	221.10
Feb.	.44	.66	−.03						S_4	−1	0	1		222.24	221.19	220.14
Mar.	.46	.66	1.00											230.87	220.00	229.13
Customer Rapport																
Nov.	.44	.00	−.04	T_A	19.64	7.11	8.02	1.40	S_1	1	1	1		8.05	8.62	9.19
Dec.	.45	.08	.04	T_B	16.21	.59	6.54	1.41	S_2	0	0	0		9.37	10.16	10.95
Jan.	.44	.33	.03	T_C	.54	3.98	−5.47	1.18	S_3	0	0	0	=	12.89	14.01	15.12
Feb.	.44	.66	−.03						S_4	−1	0	1		17.81	19.32	20.84
Mar.	.46	.66	1.00											17.52	20.27	23.03
Sales ability																
Nov.	.44	.00	−.04	T_A	65.14	11.67	.77	38.99	S_1	1	1	1		11.35	28.54	45.72
Dec.	.45	.08	.04	T_B	3.43	7.36	24.98	32.94	S_2	0	0	0		9.56	29.71	49.87
Jan.	.44	.33	.03	T_C	3.15	23.88	29.37	−.66	S_3	0	0	0	=	1.88	29.89	57.89
Feb.	.44	.66	−.03						S_4	−1	0	1		−8.08	30.83	69.75
Mar.	.44	.66	1.00											−3.64	35.38	74.39

CONCLUSIONS

The results present a clear summary of criterion variance. Evidence is found for systematic variance of persons' scores on performance measures, systematic variance of persons' scores over time, and systematic variance of scores due to individual differences. These results require a reevaluation of traditional criterion usage.

Static Dimensionality

The results indicate that performance measures taken at one time may be multidimensional. That is, the variance of any single performance measure may be represented as a sum of variance components due to the factor structure plus a component of unique variance. For example, the variance of a criterion measure in this study may be represented as the sum of four-variance components. The first component indicates variance accounted for by the Speed and Accuracy dimension; the second is an indication of the variance accounted for by the Customer Rapport dimension; the third reflects variance accounted for by the Sales Ability dimension; and the fourth-variance component reflects amount of unique variance. The percent of variance accounted for by each component is simply the squared loading on the respective component.

From this analysis it is immediately evident that in order to achieve optimum prediction, a predictor variable must be dimensionally congruent with the criterion variable. That is, both the predictor and the criterion variable must load on the same dimensions. A dimensional analysis of performance criteria and predictors allows an examination of the reasons for the correlations between predictors and criteria. The correlations are determined by shared variance accounted for by the same factors.

Dynamic Dimensionality

The consequences of orthogonal dimensions due to time are similar to those of orthogonal dimensions of performance measures. If variance of a criterion score at one point in time may be represented as the sum of orthogonal components of time variance, then optimum prediction can only be achieved if the predictor variable shares an identical factor structure due to time with the criterion variable. In the present study, both the predictor variable and the criterion variable would have to demonstrate loadings on the three time dimensions in order to achieve a maximum correlation. A dissimilar factor structure due to time between criteria and predictors will necessarily result in low predictor-criterion correlations.

Such a dimensional analysis also has implications for the classic validation study in which a predictor is correlated with a criterion at time = 1, and is assumed to be an equally good predictor of the same criterion at time = 2. From the present study it is evident that one may expect the criterion factor structure due to time to change from time = 1 to time = 2. Thus, the nature of the predictor-criterion relationship as manifest in shared variance of time components will change and the predictor-criterion correlation will probably decrease. This problem is further compounded when one realizes that the factor structure due to time of predictors is also subject to change. Therefore, the factor structure of both predictors and criteria at time = 1 may bear little resemblance to the factor structure at time = 2 and the predictor-criterion correlations at these two points in time will be the result of different configurations of shared time variance.

Individual Dimensionality

The results also indicate that individual differences variance is multidimensional and the implications parallel those above. Again, optimum prediction is achieved only if the dimensions of individual differences on the criterion variable are represented similarly on the predictor variable. Different factors accounting for individual differences in predictors and criteria will function to depress the predictor-criterion correlations. Likewise, a dimensional analysis allows the determination of the reasons for the correlations between predictors and criteria. That is, an inspection of loadings of criteria and predictors on the individual differences dimensions will demonstrate which sources of variance are shared and determine the correlation.

The Validity Problem

The dimensional complexity of criteria may very well contribute to the discouraging validities reported in studies attempting performance prediction. This study presents evidence for three multidimensional sources of variance: static dimensionality, dynamic dimensionality, and individual dimensionality. Optimum prediction of such performance measures requires predictor variables sharing a similar factor structure within each source of variance. That is, predictor variables must also demonstrate three multidimensional sources of variance (static dimensionality, dynamic dimensionality, and individual dimensionality) which are congruent with criterion sources. This conceptualization of predictor-criterion relationships argues against the use of unidimensional, highly reliable, factor pure predictor tests or batteries. Instead, optimum prediction requires multidimensional, heterogeneous tests and predictor batteries in order to adequately share variance components with the dimensionally complex criteria.

The Criterion Problem

The analysis presented allows a thorough investigation of the nature of criterion variance and suggests possible reasons for low validities. However, the problem of how to develop the single criterion score necessary for decision making remains. It has been suggested above that a single measure, measures weighted by their loadings on the first principal component, and measures predicted independently are inappropriate solutions. Instead, the investigation of performance measure dimensionality suggests two alternatives.

First, one may argue that different dimensions of performance should be used for different decisions. For example, to determine which employees would be most effective as stewardesses, one would naturally choose employees high on the Customer Rapport dimension. Or, for clerk typists, one would be more interested in persons high in Speed and Accuracy, while regional salesmen would be expected to exhibit exceptional Sales Ability. In other words, an investigator may decide *a priori* which dimension of performance is relevant to the pending decision. The predictor battery validated against the component scores on the relevant performance dimension would be used for the decision.

The second possible alternative applies the hierarchical model in a single order which has been discussed by British psychologists (Vernon, 1950). These writers assume a general factor at the top of the hierarchy below which are major group factors and minor group factors. Applying such a model to the data presented here might argue for a general factor of performance upon which all the variables load positively, and three lower order factors corresponding to the dimensions of Speed and Accuracy, Customer Rapport, and Sales Ability. The predictor battery validated against estimated factor scores on the general performance factor would be used for making all organizational decisions. The underlying assumption is that the general factor accounts for enough of the variance within each major performance dimension to permit the best decision.

Which of these two methods yields the optimum decision is open to empirical investigation. Any comparison must be made in terms of the validities of predictor batteries in predicting the criterion and the validity of the decision in achieving the desired results.

REFERENCES

Brogden, H. W., & Taylor, E. K. The dollar criterion—applying cost accounting concepts to criterion construction. *Personnel Psychology,* 1950, 3, 133–154.

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

Dunnette, M. D. A note on *the* criterion. *Journal of Applied Psychology,* 1963, **47,** 251–254.

Edgerton, H. A., & Kolbe, L. E. The method of minimum variation for the combination of criteria. *Psychometrika,* 1936, **1,** 183–187.

Ghiselli, E. E. Dimensional problems of criteria. *Journal of Applied Psychology.* 1956, **40,** 1–4.

Grant, D. L. A factor analysis of managers' ratings. *Journal of Applied Psychology,* 1955, **39,** 283–286.

Horst, P. Obtaining a composite measure from a number of different measures of the same attribute. *Psychometrika,* 1936, **1,** 53–60.

Lawler, E. E. The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology,* 1967, **51,** 369–381.

Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology,* 1963, **47,** 149–155.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika,* 1966, **31,** 279–311.

Vernon, P. E. *The structure of human abilities.* New York: Wiley, 1950.

Wiggins, N., & Hoffman, P. J. Three models of clinical judgment. *Journal of Abnormal Psychology,* 1968, **73,** 70–77.

Received: April 5, 1971