

# Discriminability in Multidimensional Performance Evaluations

**Ditsa Kafry**

**University of California, Berkeley**

**Rick Jacobs**

**University of California, Riverside**

**Sheldon Zedeck**

**University of California, Berkeley**

A series of behavioral expectation scale (BES) applications were analyzed in an attempt to point out an appropriate number of dimensions to be included in such studies. Data from 4 independent samples, representing 3 different occupations, and incorporating a total of 436 multidimensional evaluations were subjected to factor analysis. Results reflected the lack of unique information contributed when the number of dimensions exceeds 9. The problem of lack of dimension independence was discussed in terms of theory and application to multidimensional performance evaluation. Suggestions are advanced for limiting the number of dimensions as a potential solution to information redundancy.

Behavioral expectation scales (BES) as a means of evaluating workers' performance have been the subject of many recent research studies. A series of these investigations have tested the independence of performance evaluation scales (Dickinson & Tice, 1973; Nealey & Owen, 1970; Zedeck & Baker, 1972). These studies found that, in general, the scales demonstrate a certain degree of convergent validity, but lack discriminant validity. The lack of discriminant validity can be seen as a consequence of high dimension intercorrelations and the inclusion of redundant information. Additional

investigations have tested the impact of rater participation in scale construction on dimension independence. Friedman and Cornelius (1976) have shown that participation in scale development leads to an increase in convergent validity, but has little effect on discriminant validity. Bernardin, LaShells, Smith, and Alvares (1976) reported that participants in scale construction showed no increase in discriminability over non-participants. These results imply that while participation in BES development may increase agreement among raters, it does not improve raters' ability to discriminate among the dimensions.

One potential explanation for the lack of independence in ratings might be the limit of an individual's ability to cognitively process information on several diverse dimensions, irrespective of participation in scale development. Thus, the raters' inability to handle the many discriminations required for performance appraisal results in assigning evaluations to all the dimensions based on the limited information available. An alternative explanation may be that there is a "true" relationship among performance dimensions or subsets of dimensions. It may be that for a given job, conceptually pure dimensions represent different elements of a more general factor. With this true underlying clustering of dimensions, it is not surprising that scale evaluations yield high degrees of intercorrelation. This latter

explanation does not readily lend itself to empirical validation, since "true" values, in a performance evaluation sense, are not known. Specification of "true" scores might be possible under controlled laboratory settings and should be the subject of future research in this area.

The limit of an individual's capacity to process information suggests that cognitive constraints might be related to the number of evaluation dimensions used for performance assessment. It may be that the greater the number of dimensions, the more difficult the discriminability. Previous BES studies have either generated or employed between 3 (Dickinson & Tice, 1973) and 24 (Zedeck & Kafry, 1977a) evaluation dimensions. However, none of the papers discussed information redundancy or the appropriate number of rating dimensions, appropriate in terms of representing unique information not included in other dimensions. An answer to the question concerning nonredundant dimensions has theoretical, as well as applied, value. The theoretical value is relevant to understanding the cognitive information-processing capabilities of the raters. The relevant task is to discover how many conceptually independent evaluations people are capable of performing. On the applied level, discovering the appropriate number of dimensions may reduce the amount of time and energy required for proper use of BES. According to BES procedures, the raters are tasked with observing the behaviors of several subordinates, comparing these behaviors to scale anchors, and providing a numerical evaluation together with an example of actual behavior for each subordinate on each dimension. The "tremendous amount of effort" (Campbell, Dunnette, Arvey, & Hellervik, 1973) might be reduced once several dimensions are either combined with other dimensions or eliminated.

The study reported here is designed to offer information on the appropriate number of dimensions included in BES applications. The criterion for determining this appropriate number is one of nonredundant dimension information.

The main focus of this investigation is the reduction of the number of original dimensions (rated dimensions) while still retaining a high percentage of original information.

### Method

Data were gathered from four independent applications of BES. For each study, the basic input variables were the ratings for each subject on each BES dimension. The number of dimensions ranged from 9 (Study A) to 24 (Study D). The ratio of number of subjects relative to number of input variables is always a consideration for interpretation of factor analytic work. Within this study, the samples surveyed varied from a ratio of approximately 20 to 1 (Study A) to a ratio of 3 to 1 (Study C). Care should be taken in interpreting the results as a function of this small subjects to variables ratio, and replication with larger samples would certainly be important. The BES assessments are described in detail below.

#### Study A

One hundred seventy-seven students from three statistics courses rated their professors by using a set of BES constructed in an earlier study. These scales (developed by Harari & Zedeck, 1973) yield 9 scores on the behavioral dimensions of Depth of Knowledge, Delivery, Organization, Interpersonal Relations with Students, Relevance, Testing, Grading, Assignments and Work Load, and Inspiration and Motivation. Following a brief session to familiarize the students with the scales, numerical evaluations were completed during the last 15 minutes of regularly scheduled classes.

#### Study B

Police officers were evaluated by their immediate supervisors, police sergeants. Ninety-five officers were rated on an 11-dimension BES format created for the study. The dimensions rated

were Beat Attention and Public Contact, Communication Skills, Decision Making and Initiative, Driving Skills, Investigations, Knowledge of Laws, Physical Fitness and Appearance, Relations with Coworkers, Relations with Supervisors, Roll Call and Briefing, and Safety Practices. The evaluations were part of a larger project designed to study the relationships of performance variables with other organizational outcome measures.

### Study C

Forty-nine public health nurses were evaluated by their supervisors. Evaluations were performed using a 16-dimension BES developed by the supervisors and containing the following job performance categories: Assessment of Family Problems, Collaboration with Others, Dependability, Flexibility, Initiative and Involvement, Relationships with Co-workers, Relationships with Supervisors, Interviewing Skills, Knowledge, Objectivity, Organizational Ability, Professional Growth, Self-Awareness, Stability, Teaching Skills, and Written Communications.

### Study D

One hundred fifteen public health nurses from two counties, independent of Study C, were evaluated by their supervisors on a 24-dimension BES which was developed for this study and was part of a larger project described elsewhere (Zedeck & Kafry, 1977a, 1977b). The BES used contained the 16 performance dimensions listed in Study C plus the following: Advocacy; Empathy, Interest, and Relations with Clients; Follow-up Actions; Integrity; Leadership Potential; Patience; Planning and Recommending; and System Orientation.

### Analysis

The evaluation data from each BES study were subjected to factor analysis. In each study the correlation matrix of dimension ratings was

generated and analyzed with squared multiple correlations as communality estimates and with the stipulation that the number of factors extracted equal the number of input dimensions. Each solution was then rotated to simple structure via the varimax procedure.<sup>1</sup> The results of these procedures are summarized in Table 1 and point to highly consistent findings.

### Results and Discussion

Most pertinent to this investigation, the data support the conclusion that 9 or fewer factors accounted for 90% or more of the total input variability in each sample. The analyses revealed that dimensions can be reduced to factors at a rate of between 56% (from 9 dimensions to 4 factors for Sample A) to 63% (from 16 dimensions to 6 factors for Sample C and from 24 dimensions to 9 factors for Sample D). These results, consistent with previous investigations, depict a lack of discriminant validity and point up two major problems with BES methodology and multidimensional performance appraisal in general. Of greatest concern is what happens between dimension generation and dimension utilization. During developmental stages, potential raters are asked to assign behaviors to specific job dimensions. This retranslation process requires participants to distinguish conceptually between the different areas of job performance. In the development of the 4 sets of scales described in this study, raters were able to assign the many behavioral examples to from 9 to 24 scales. However, in the use of these scales, the conceptually distinct dimensions demonstrated high degrees of interdependence. The results of the factor analyses show that 4 to 9 factors accounted for approximately 90% or more of the total dimension variance. This finding is consistent with other studies which found no more than 7 nontrivial factors (Campbell et al., 1973; Keaveny & McGann, 1975). Addition-

---

<sup>1</sup>Complete information regarding results from the factor analysis is available from the second author.

Table 1  
Summary of Results from Studies A through D

Variable	Sample			
	A University Students	B Police Officers	C Public Health Nurses	C Public Health Nurses
Number of Evaluations	177	95	49	115
Number of Dimensions in Original BES	9	11	16	24
Number of Factors	4	4	6	9
Percent of Var- iation in Orig- inal Dimensions Represented af- ter Reduction	93.8	90.5	90.4	91.0

ally, it should be noted that Zedeck and Kafry (1977b) reported that in general, 4 elements or less accounted for most of the predictable variance in an individual's judgments of another's overall performance.

It is unclear whether the high scale intercorrelations are a function of "true" relationships among performance dimensions or whether they are the result of cognitive overload. The first explanation leads to a prediction that the structure of intercorrelations would be replicated for future samples evaluated on the same scales. The second explanation predicts that either lack of discriminability will be found whenever the total number of dimensions used exceeds 9, regardless of the sample or the dimensions used, or that dimension independence will be a function of individual differences in level of cognitive complexity of the rater. Schneier (1977) recently addressed this issue and found that cognitively complex raters showed less halo and suggested that they may be able to better differentiate among dimensions. The choice between the

"true" relationship and the "overload" explanation requires further research. However, regardless of the correct answer, it behooves scale developers to limit the number of rated dimensions to be consistent with both the minimum number requisite for adequate job description and the maximum number of discriminable stimuli for a group of raters.

The appropriate number of rated dimensions is an applied problem for BES and other performance appraisals systems. Given that individuals may only be capable of evaluating a limited number of dimensions, there is minimal value in requiring discriminations beyond this number. The applied nature of this problem of discriminability mandates immediate suggestions for the reduction of the number of dimensions used in the BES procedure. Results from this study and similar investigations provide a partial solution to the problem of appropriate number of dimensions by indicating the minimum number of factors required to account for overall scale variation. Unfortunately, the di-

rect results from factor analysis, factor scores, do not answer the problem completely. First, using factor scores still requires evaluators to provide ratings on all dimensions, a task which is both time consuming and lacking in discriminant validity. Of even greater importance is the fact that factor scores are often relatively uninterpretable in terms of what is being evaluated. One method for understanding the performance dimensions represented by a factor is to examine the matrix of factor loadings. Dimensions with high factor loadings could be used to replace the factor. A criterion for replacement could be to choose the dimension or dimensions with the highest factor loading within a given factor. This solution would require that the full set of dimensions be rated at least once to provide the necessary data.

Other, nonstatistical, solutions to the problem of requesting raters to assign evaluations to more than nine dimensions are also possible. One potential procedure relies on the judgments of the participants in scale development. The suggestion here is to incorporate in the scales only those dimensions which developers feel are performed often and/or are important for successful execution of job responsibilities. This solution would require each BES participant to rate each dimension on these two variables during scale construction. Upon analysis, a composite of the most often performed and the most important scale dimensions would appear on the final scales.

A second potential solution could be to use different subsets of scales over time. By dividing the total number of dimensions into more manageable subsets, it may be that evaluators could then make necessary dimension discriminations. Such a solution might be preferred if raters are uncomfortable over rating ratees on only a few dimensions when the raters feel that there are many more conceptually distinct dimensions. That is, raters may feel the appraisal instrument is deficient regardless of the statistical superiority of the scales. This would be especially true in circumstances where performance evaluation

is coupled with feedback and the ratings are used as a means of improving employee performance across the range of job dimensions. This solution could be operationalized by requiring each rater to use a certain subset of dimensions within a specified evaluation time period until all dimensions are rated over a longer evaluation period. The selection of dimensions could be done randomly or could be based on judgmental decisions of dimension relevance for certain periods of work performance (e.g., probation period, training for a specialty, training for a deficiency).

It appears from the data in this study and in similar studies using comparable methodology that the number of evaluative dimensions can be statistically reduced via factor analysis. Such solutions, although informative with respect to scale information redundancies, do not provide necessary results for appropriate dimension utilization. Suggestions were presented to help minimize the problem of lack of dimension discriminability in multidimensional performance evaluations.

## References

- Bernardin, H. J., La Shells, M. B., Smith, P. C., & Alvares, K. M. Behavioral expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology*, 1976, 61, 75-79.
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervick, L. V. The development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, 1973, 58, 15-22.
- Dickinson, T. L., & Tice, T. E. A multitrait-multimethod analysis of scales developed by retranslation. *Organizational Behavior and Human Performance*, 1973, 9, 421-438.
- Friedman, B. A., & Cornelius III, E. T. Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. *Journal of Applied Psychology*, 1976, 61, 210-216.
- Harari, O., & Zedeck, S. Development of behaviorally anchored scales for the evaluation of faculty teaching. *Journal of Applied Psychology*, 1973, 58, 261-265.
- Keaveny, T. J., & McGann, A. F. A comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, 1975, 60, 695-703.

- Nealey, S. M., & Owen, T. W. A multitrait-multi-method analysis of predictors and criteria of nursing performance. *Organizational Behavior and Human Performance*, 1970, 5, 348–365.
- Schneier, C. E. Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. *Journal of Applied Psychology*, 1977, 62, 541–548.
- Zedeck, S., & Baker, H. T. Nursing performance as measured by behavioral expectation scales: A multitrait-multirater analysis. *Organizational Behavior and Human Performance*, 1972, 7, 457–466.
- Zedeck, S., & Kafry, D. Evaluation of the developers of behavioral expectation evaluation scales. *JSAS Catalogue of Selected Documents in Psychology*, 1977, 7, 123. (NS.1618) (a)
- Zedeck, S., & Kafry, D. Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance*, 1977, 18, 269–294. (b)

### Acknowledgments

*This research was partially supported by an inter-campus research opportunity grant from the University of California, Riverside.*

### Author's Address

Send requests for reprints or further information to Rick Jacobs, Department of Psychology, University of California, Riverside, Riverside, CA 92521.