

## Brief Report

# Behavioral Expectation Scales versus Nonanchored and Trait Rating Systems: A Sales Personnel Application

**John M. Ivancevich**  
**University of Houston**

There are presently available several empirical comparisons between behavioral expectation scales (BES) and other rating scales (Bernardin, 1977; Borman & Dunnette, 1975; Burnaska & Hollman, 1974; Keaveny & McGann, 1975). In many of these studies the rater-ratee population has consisted of faculty members and students (Bernardin, 1977; Keaveny & McGann, 1975; Schwab, Heneman, & Decotiis, 1975). Only a handful of scientifically sound investigations comparing BES and other rating scales have used manager-subordinate populations.

The importance of sales personnel in organizations, the general lack of previous research using sales employees as subjects in examining BES, and the significance of performance evaluation prompted the present study. Specifically, the study examined estimates of rating leniency, halo error, interrater agreement, and the degree of ratee differentiation of BES, nonanchored, and trait evaluation systems.

### Previous Studies and The Present Study

A well-designed study by Borman and Dunnette (1975) investigated Navy officer raters who used BES, scales containing the same definitions but without the behavioral anchors used in the BES, and a series of trait dimensions. The BES had reduced leniency and halo error and restriction of range of ratings.

Field sales personnel have rarely been used as subjects in developing rating systems that have some practical value for both the rater and ratee. Previous research comparing BES to other evaluation systems led to the development of the hypothesis that the use of BES would result in less psychometric error than the nonanchored and trait systems. The psychometric considerations in this study were leniency error (a shift in mean ratings in a favorable direction), halo error (pooled variance across performance dimensions), interrater reliability (the degree of agreement between raters on the dimensions for each ratee), and level of ratee discrimination (difference in standard deviations of ratings on each dimension, with higher standard deviations representing greater discrimination).

### Method

The present study was conducted in a medium-sized organization that conducts business in urban areas, primarily on the East coast and in the Midwest. The managers in the company had been

---

**APPLIED PSYCHOLOGICAL MEASUREMENT**  
**Vol. 4, No. 1 Winter 1980 pp. 131-133**  
**© Copyright 1980 West Publishing Co.**

rating sales personnel by employing a 10-trait (10-point scale) performance appraisal for the past 30 months. The company expressed an interest in instituting a new rating system because of numerous rater and ratee complaints about the present system.

Three rating scales (a BES, a nonanchored scale using the same performance dimensions as the BES, and a trait scale) were used by the raters (executives and district managers) to rate the performance of the subordinates. In a manner similar to that reported by Borman and Dunnette (1975) the raters evaluated the ratees in 2 sessions that were 30 days apart. The rating scale assignments to the 19 raters were reversed for the 2 separate rating sessions. One group contained 3 executive officers and 8 district managers, who rated 48 subordinates; the other group had 2 executive officers and 6 district managers, who rated 43 subordinates.

### Data Analyses and Results

Four separate analysis of variance models were employed to examine the data. The dependent variables were (1) the grand means for each rating scale, the leniency error measure; (2) standard deviations for the measure of the ability of a rater to differentiate scale dimensions; (3) pooled variance across performance dimensions, the halo measure; and (4) interrater reliability, the measure of agreement among raters.

The analysis of variance resulted in only one significant ( $p < .01$ ) finding for the format main effect. The data analysis also revealed that the mean rating for the BES was lower than the nonanchored and trait rating system means. However, the means of the BES and the nonanchored rating, when analyzed together, did not statistically differ from the trait mean. It was also determined that the differences between rating system standard deviations were small.

The data indicated that the BES was slightly better than either of the other two formats. The mean halo score for the BES was significantly better (less halo error) than the corresponding mean for the nonanchored scale. Interrater agreement results revealed that the BES was modestly better than the nonanchored scale ( $p < .05$ ). There was no significant difference found in interrater agreement when trait versus BES and nonanchored scales were compared.

The overall results suggest that portions of the hypothesis can be accepted for the present study and sample. Specifically, the BES was only modestly superior to the nonanchored and trait rating system with regard to minimizing halo error and improving interrater agreement.

### Discussion

This study, similar to Borman and Dunnette's (1975), showed that the mean differences among the dependent variables for the three rating systems were small in relation to the magnitude of the standard deviations for each of the dependent variables. The range was from about one-half of a standard deviation for the mean variables to approximately one-fifth of a standard deviation for the other three dependent variables.

The issue of the potential benefits of utilizing the BES should be faced by any organization considering their possible use. The intuitive appeal of the BES system for organizations may evolve from the value of clearly defining job performance dimensions and allowing subordinates to participate in scale development. This benefit (Blood, 1974) may provide some intrinsic value to the personnel participating in the development of the BES.

The results from this study were similar to those found by Borman and Dunnette (1975) using U.S. Navy personnel. In the present study, sales personnel, who performed quite different job tasks

than U.S. Navy officers, served as subjects. Thus, the replication of the Borman and Dunnette (1975) findings in a nonmilitary setting and with a sales personnel group suggests that the usefulness of BES over other rating formats probably will not be found in psychometric superiority. Psychometric errors seemed to remain or to continue to exist no matter how well developed the format; and to justify the time and effort expended on developing behavior-based evaluation scales, other advantages from using BES need to be found.

### References

- Bernardin, H. J. Behavioral expectation scales versus summated scales: A fairer comparison. *Journal of Applied Psychology*, 1977, 62, 422-427.
- Blood, M. R. Spin-offs from behavioral expectation scale procedures. *Journal of Applied Psychology*, 1974, 59, 513-515.
- Borman, W. C., & Dunnette, M. Behavior-based versus task-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, 1975, 60, 561-565.
- Burnaska, R. F., & Holliman, T. D. An empirical comparison of the relative effects of rater response biases on three rating scale formats. *Journal of Applied Psychology*, 1974, 59, 307-312.
- Keaveny, T. J., & McGann, A. F. Comparison of behavioral expectation scales and graphic rating scales. *Journal of Applied Psychology*, 1975, 60, 695-703.
- Schwab, D. P., Heneman, H. C., & DeCotiis, T. Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology*, 1975, 28, 549-562.