Effects of Performance Appraisal Format on Perceived Goal Characteristics, Appraisal Process Satisfaction, and Changes in Rated Job Performance: A Field Experiment

AHARON TZINER  
Department of Psychology  
University of Montreal, Canada

RICHARD E. KOPELMAN  
Management Department  
Baruch College, CUNY

NEOMI LIVNEH  
Tel Aviv Medical Center, Israel

ABSTRACT. We hypothesized that a behavioral observation scale (BOS-based) performance appraisal and review process, compared with a graphic rating scale (GRS-based) process, would yield higher levels of goal clarity, acceptance, and commitment; appraisal process satisfaction; and improvement in job performance. Data from a field experiment involving 16 managers and 115 subordinates confirmed all three predictions. These results, and past research, suggest that the primary advantage of a BOS-based performance appraisal may not be psychometric, but may instead be in terms of its effects on work behavior and job performance.

A MAJOR PURPOSE OF PERFORMANCE APPRAISAL is to modify work behavior (Austin, Villanova, Kane, & Bernardin, 1991; Bernardin & Beatty, 1984; Murphy & Cleveland, 1991). Although some theorists hold that individual behavior tends to be highly repetitive and impervious to outside influences (Campbell & Pritchard, 1976; Lewin, 1938), basic to any performance appraisal process is the assumption that accurate and timely feedback can change behavior. Yet, Nathan, Mohrman, and Milliman (1991, pp. 352–353) have cautioned that “whether performance appraisal reviews actually change
subsequent employee performance and work attitudes is still very much an open question."

One aspect of performance appraisal that has received considerable attention over the past two decades is the rating format (Feldman, 1981; Tziner, 1990). In particular, much effort has been devoted to comparing the psychometric properties of behavior-based scales versus graphic rating scales, with the result that the former has typically demonstrated smaller advantages than anticipated (Landy & Farr, 1980, 1983). Yet, no researchers have examined what for many applied psychologists is a more crucial issue: Do behavioral (in comparison to graphic) appraisal formats yield superior results in terms of employee performance and work attitudes?

We assumed that a behavioral observation scale (BOS-based) performance appraisal and review process should lead to higher levels of performance and satisfaction with the review process than would a graphic rating scale (GRS-based) approach for three reasons. First, the BOS-based process would convey more clearly exactly what an individual should do. Rather than enjoining the person to "improve your sense of responsibility," and to "sharpen your leadership skills," behavior-based feedback "pinpoints" (Miller, 1978) the specific desired/undesired actions to be taken/avoided (Latham & Wexley, 1981). Second, BOS-based feedback would be generally more acceptable (in comparison to GRS-based feedback) because it is seen as more factual, objective, and unbiased (Kopelman, 1986). Third, BOS-based feedback would be more conducive to setting performance goals that are specific rather than vague and job-related rather than nonrelated. These goal properties (specificity and relevance) would likely influence the extent to which appraisees perceive goals to be clear, accept them, and commit to accomplishing them (Carroll & Schneier, 1982; Dolan & Schuler, 1987; Huber, 1985; Pinder, 1984).

In the present research, we examined how the performance appraisal format (BOS vs. GRS) influences perceived goal characteristics (e.g., goal clarity, goal acceptance, and goal commitment), process satisfaction (i.e., attitudes regarding the performance appraisal and review process), and job performance.

In one prior field study, Tziner and Kopelman (1988) found that a BOS-based performance appraisal and review process led to higher levels of goal clarity, acceptance, and commitment than did a GRS-based approach; however, no data were examined pertinent to employee affect or job performance. To our knowledge, the present research is the first attempt to compare directly the effects of BOS- and GRS-based performance feedback on employee job

Address correspondence to Richard E. Kopelman, Management Department, Box 507, Baruch College, City University of New York, 17 Lexington Avenue, New York, NY 10010.
performance and affective reactions to the review process in a work organization. In addition to replicating the earlier research in a different setting (a hospital vs. an airport authority), we examined longitudinal as well as concurrent data and formulated the following three hypotheses:

1. In comparison with a GRS-based performance appraisal and review, a BOS-based process will lead to higher levels of goal clarity, acceptance, and commitment.
2. In comparison with a GRS-based performance appraisal and review, a BOS-based process will lead to higher levels of satisfaction with the appraisal process.
3. In comparison with a GRS-based performance appraisal and review, a BOS-based process will lead to higher levels of subsequent job performance ratings.

Method

Participants

Participants were 16 nurse-managers and their 115 nurse-subordinates. The managers (raters) were randomly assigned to one of two groups, the BOS or the GRS performance appraisal group. Of the subordinates (ratees), 53 were assigned to the BOS group, 62 to the GRS group. All 16 raters were women, ranging in age from 29 to 58 years (\(M = 45.75; SD = 8.74\)) and in level of education from 15 to 21 years of schooling (\(M = 16.62; SD = 1.43\)).

Procedure

The experiment was conducted at a large hospital complex in the Tel Aviv area. The 16 raters received performance appraisal and feedback training in both formats prior to the beginning of the appraisal and review process. The objectives of the training program were derived from the following model conceived by Borman (1982):

1. Develop rater skill in observing and recording only performance-related behaviors.
2. Make raters aware of the distinction between performance categories (dimensions) and component behaviors.
3. Train raters in setting specific and concrete goals that correspond to the performance dimensions under review.
4. Foster a shared understanding among raters as to the content and meaning of each performance dimension.

The training also attempted to develop in raters the ability to provide explicit, factual, and candid performance feedback.
Training methods included lectures, discussions, videotapes of models (actual nurses), role-playing exercises, practice, and constructive feedback. Emphasis was placed on practice and feedback, as these components have been found most effective in creating lasting effects on rater performance (Bernardin & Beatty, 1984). The 1-day training program provided practice in observing, recognizing, and recording performance-related behaviors and in formulating evaluative judgments using a diary and a prerecorded videotape of models. Role-playing exercises focused on how the rater should properly conduct the performance feedback session and formulate performance goals.

Subsequent to training, performance appraisals were conducted using the two rating formats, BOS and GRS. The BOS format was developed along the lines described by Latham and Wexley (1981). Each behavioral item was scored on a 6-point scale indicating the frequency with which a ratee exhibited the particular behavior. Endpoints ranged from almost never (1) to almost always (6). The BOS form consisted of 85 behavioral items classified into 12 performance dimensions derived from 250 items generated by 10 different nurse-managers who did not otherwise participate in this study. Items were retained only if assigned to the same performance dimension by at least 80% of these raters; dimensions were retained only if deemed important by at least 70% of the raters.

The GRS appraisal instrument and the BOS form contained the same 12 performance dimensions (also labeled identically): organizing and managing, initiative, responsibility, professional ethics, discipline, orderliness, orientation/information provided to patients and families, application of professional knowledge, attitude toward patients, cooperation with peers, organizational commitment, and performance under stress. The GRS appraisal also used a 6-point scale, with the endpoints of very untypical (1) and very typical (6). Both forms also provided identical narrative definitions of each performance dimension. Thus, the two rating formats were identical with respect to performance dimensions, nominal descriptions and conceptual definitions, and response ranges. Differences in instrument length were controlled by comparing the means of the 12 dimension scores for both formats.

Design

After completing the training workshop, the raters were randomly assigned to one of the two experimental conditions: BOS-based or GRS-based performance appraisal and review. We used a random assignment to ensure (to the extent possible) equivalence of the two groups in terms of all potentially relevant individual difference variables. The trainees were not informed about the existence of two formats; nor did the trainer, Naomi Livneh, know who would be assigned to which group.
In both groups, performance ratings were communicated and discussed with the ratees. Toward the latter part of that session, goals were formulated that corresponded to the feedback, in that they were the areas of performance that particularly needed improvement. Raters indicated that progress in achieving these goals would be reviewed at the next performance review session. Rates were told that these reviews would be used to improve individual (and organizational) performance but not for administrative (e.g., salary allocation) purposes. Examples of goals set after GRS- and BOS-based performance reviews are shown in the Appendix.

Shortly after the conclusion of the review (within 48 hours), a questionnaire was administered to ratees, not in the presence of the raters. We asked questions pertinent to goal clarity, goal acceptance, goal commitment, and ratee satisfaction with the performance review.

Eight months after the first review, another wave of appraisals was conducted, using precisely the same forms, performance review method, questionnaires, and following the same experimental procedure. Unfortunately, longitudinal panel data could be matched only with respect to job performance; for the other two dependent variables, individual data could not be matched across time periods. This precluded the use of confirmatory factor analysis to examine comprehensively the validity of the causal model proposed here.

Measures

Goal acceptance was measured with the same two items as in a previous study (Tziner & Kopelman, 1988): To what extent do you accept the goals you have been assigned? To what extent do you agree that the performance goals set up for you are reasonable? Responses on a 6-point Likert scale ranged from to a very limited extent (1) to to a very great extent (6). These items were combined into a single measure, with an internal consistency reliability estimate ($\alpha$) of .77 (Time 1) and .86 (Time 2). (In the previous study, alpha was .70.)

Goal commitment was assessed using two items: To what extent are you determined to achieve the performance goals? To what extent are the performance goals important to you? Responses were provided on the same 6-point Likert scale used to measure goal acceptance. The two items were combined into a single score with an internal consistency reliability estimate ($\alpha$) of .91 (Time 1) and .88 (Time 2). (In the prior study, alpha was .84.)

Goal clarity was measured using four items that inquired about the extent to which (a) it is clear what course of action you should take in order to accomplish your performance goals, (b) the information provided to you by your supervisor during the performance review will help you attain your
goals, (c) the information provided to you by your supervisor during your performance review was sufficiently lucid, and (d) the information provided to you by your supervisor during your performance review was sufficiently detailed. Responses to these items were provided on the aforementioned 6-point Likert scale. A single score was computed with an estimated internal consistency (α) of .85 (Time 1) and .87 (Time 2). (In the prior study, alpha was .82.)

Ratee satisfaction with the performance review process was assessed using the 8-item instrument developed by Russell and Goode (1988). Sample items included: (a) Throughout the performance review I felt that I had the possibility to frankly discuss work-related issues with the supervisor; (b) I felt that the performance review was unfair; (c) The performance review helped me better understand how to perform my job; (d) I was satisfied with my performance review. The ratees were requested to report the extent to which they endorsed each of these statements using the aforementioned 6-point Likert scale. Responses were combined into a single measure with an internal consistency reliability estimate (α) of .65 (Time 1) and .70 (Time 2).

BOS-based performance was measured by summing the 12 dimension scores. A 6-point scale was used with endpoints of almost never (1) and almost always (6). Internal consistency reliabilities (alphas) for the BOS-based performance dimensions ranged from .46 to .95 (Time 1) and .45 to .94 (Time 2); the median alphas, respectively, were .79 and .80. For the summed measures, alphas were .91 and .93, respectively. GRS-based performance was assessed by summing the 12 dimension scores. Median intercorrelations (r) across dimensions were .42 (Time 1) and .47 (Time 2). The Cronbach alphas were .89 and .91, respectively.

Results

A preliminary analysis of the data indicated an average intercorrelation of .63 among the three perceived goal characteristics (clarity, commitment, and acceptance). We then examined the three measures using a two-factorial multivariate analysis of variance (MANOVA), with time and format as the factors. The results of this analysis indicated that there were significant differences in perceived goal characteristics across the two experimental groups: $F(3, 111) = 5.44, p < .001$, Wilks’s $\lambda = 0.92$. Further, the effects of time and the interaction of format and time were statistically insignificant: $F(3, 111) = 1.35, p < .26$, Wilks’s $\lambda = 0.98$, and $F(3, 111) = 0.14, p < .38$. Wilks’s $\lambda = 0.98$, respectively.

We conducted a two-way analysis of variance (ANOVA) to understand more thoroughly the effect of format (Table 1). Again, neither time, nor the Format × Time interaction, was significant. We also conducted a priori t tests
(Wildt & Ahtola, 1978, p. 30) to compare mean levels of goal clarity, goal acceptance, and goal commitment across the two performance appraisal formats. As hypothesized, the BOS format yielded higher levels of each criterion at Time 1 (Table 2). We also found that BOS-based performance appraisal and review yielded higher levels of goal clarity, goal acceptance, and goal commitment at Time 2 (Table 3).

As predicted, the BOS-based performance appraisal and review yielded a higher level of rater satisfaction than did the GRS-based process (Hypothesis 2). Mean process satisfaction scores at Time 1 were 4.43 and 4.01, respectively, \( t(101) = 2.76, p < .01 \); at Time 2 the mean scores were, respectively, 4.48 and 4.08, \( t(107) = 2.82, p < .01 \).

With respect to the effect on rated performance (Hypothesis 3), we found that mean performance scores were essentially the same in the GRS and BOS conditions at Time 1: 5.11 and 5.10, respectively. But whereas there was a nonsignificant change in rated performance in the GRS condition, to 5.13, \( t(106) = .71 \), overall performance in the BOS condition improved modestly, to 5.22, \( t(104) = 1.12, p = .13 \). More important, examination of the change in individual performance indicated a significant improvement in the BOS condition, \( t(52) = 2.77, p < .01 \), and only an insignificant improvement in the GRS condition, \( t(53) = .71 \).

**Discussion**

Our three hypotheses were supported. As predicted, the BOS-based process of performance appraisal and review produced consistently higher levels of goal clarity, acceptance, and commitment compared with the GRS-based approach. Goal clarity was enhanced in the BOS-based performance review by setting goals that related to specific behaviors (e.g., will be capable, by next appraisal, of fully operating the respiratory machine located in her department), thus pinpointing the precise course of action needed to accomplish the set goals. Moreover, as has frequently been suggested, the specificity of a performance review may affect acceptance of set goals. Also, when raters
TABLE 2
Effect of Appraisal Format on Three Goal Characteristics: Time 1

<table>
<thead>
<tr>
<th>Characteristic/format</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>43</td>
<td>5.44</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>63</td>
<td>5.12</td>
<td>0.88</td>
<td>1.88*</td>
</tr>
<tr>
<td>Acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>43</td>
<td>5.47</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>63</td>
<td>4.94</td>
<td>1.19</td>
<td>2.64**</td>
</tr>
<tr>
<td>Commitment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>43</td>
<td>5.49</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>63</td>
<td>5.13</td>
<td>1.20</td>
<td>1.76*</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.

TABLE 3
Effect of Appraisal Format on Three Goal Characteristics: Time 2

<table>
<thead>
<tr>
<th>Characteristic/format</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>51</td>
<td>5.50</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>45</td>
<td>5.07</td>
<td>0.87</td>
<td>2.65**</td>
</tr>
<tr>
<td>Acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>51</td>
<td>5.28</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>45</td>
<td>4.64</td>
<td>1.38</td>
<td>2.82**</td>
</tr>
<tr>
<td>Commitment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS</td>
<td>51</td>
<td>5.31</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>GRS</td>
<td>45</td>
<td>4.96</td>
<td>1.18</td>
<td>1.84*</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.

review progress against specific job-related goals using a documented process of performance review, this is likely to enhance commitment to goals.

In contrast, a GRS-based performance review is prone toward yielding generic goals (e.g., will broaden her professional knowledge). Thus, the ratee is likely to experience fuzziness regarding the activities to pursue in order to accomplish the set goals. And if goal clarity is lower than for a BOS-based...
review, goal acceptance and commitment will also tend to be lower. We believe this line of reasoning accounts for the demonstrated superiority of the BOS over the GRS-based performance review in terms of goal clarity, acceptance, and commitment.

The BOS-based process also produced higher levels of satisfaction with the appraisal process. Mount (1983) has previously reported that employee appraisal satisfaction is related to a general satisfaction factor, which includes affect toward the appraisal system, the quality of performance feedback, and the extent to which the appraisal form aids discussion of performance and how it facilitates the formulation of personal development plans. In light of the behavioral specificity of the BOS-based performance review, it follows that this format should foster the provision of credible feedback and the setting of acceptable goals and that, in the present study, a BOS-based performance review yielded higher satisfaction with the process than the GRS-based performance review did.

Finally, there was no change in rated performance using the GRS-based approach, but performance improved significantly using the BOS-based process (on a within-person basis). Given the widespread use of GRS-based performance appraisals (Teel, 1980), the present results underscore the observation by Nathan et al. (1991) that the typical performance appraisal review may not change performance. Yet the BOS-based approach did lead to significantly (albeit modestly) improved job performance.

Clearly, future researchers should obtain data that are amenable to confirmatory factor analysis. Further, objective data on such goal properties as specificity and job-relatedness should be obtained using the reliable external assessment of goal statements. Such information would help in confirming the causal model of linkages between goal properties, perceived goal characteristics, process satisfaction, and improved job performance.

Although we did not examine a causal model here, we believe that the current research has brought us closer to understanding the processes triggered by the two rating formats (BOS and GRS), an issue that, according to Murphy and Cleveland (1991), has largely been ignored. Notwithstanding decades of research on the psychometric properties of these (and other) rating formats, the real payoff may be in terms of changes in on-the-job work behavior and job performance.

APPENDIX
Examples of Goals Derived From GRS- and BOS-Based Performance Appraisal and Review Processes

BOS-Based Goal Statements

Will compile, for presentation at the next professional staff meeting, materials concerning how to prepare a patient scheduled to undergo cataract surgery.
APPENDIX (continued)

Will be capable, at the next appraisal, of fully operating the respiratory machine located in her department.

Will attend a workshop aimed at increasing her sensitivity to the patient’s family concerns.

GRS-Based Goal Statements

Will demonstrate more involvement at work.

Will broaden her professional knowledge.

Will take steps to prepare herself for promotion to a job of higher responsibility.

REFERENCES


*Received October 13, 1992*