Is there a Preferred Performance Rating Format?  
A Non-psychometric Perspective

Aharon Tziner*  
School of Business Administration, Netanya University College, Israel

Richard E. Kopelman  
Zicklin School of Business, Baruch College, USA

Les échelles d’appréciation constituent une forme d’évaluation des performances qui a rencontré un grand succès ces dernières décennies. En particulier, de gros efforts ont été consacrés au développement d’échelles relativement insensibles aux erreurs et biais cognitifs. Mais on s’est beaucoup moins intéressé au fait de savoir si et comment le type d’évaluation affecte les attitudes professionnelles et les réactions des personnes évaluées. Des données en provenance de quatre études différentes avec des échantillons tirés dans deux pays (Israël et le Canada) apportent des éléments en faveur de l’idée selon laquelle une évaluation de la performance basée sur le BOS peut être supérieure aux autres méthodes en ce sens qu’elle entraîne des conséquences plus positives au niveau des attitudes.

One aspect of performance appraisal that has received considerable attention over past decades is the rating format. In particular, much effort has been devoted to developing rating scales that are relatively impervious to cognitive rating errors and biases. However, much less attention has been accorded the issues of whether and how an appraisal’s format affects work attitudes and reactions of ratees. Data collected in four separate studies and with samples in two nations (Israel, Canada) lend credence to the proposition that a BOS-based performance appraisal and review may be superior to other appraisal methods in terms of yielding more favorable attitudinal effects.

INTRODUCTION

For more than twenty years, research comparing the psychometric properties of behavior-based versus graphic rating scales has demonstrated some advantages of the former, although the effects have been less than anticipated (Landy & Farr, 1980). Specifically, these original investigations

* Address for correspondence: School of Business Administration, Netanya University College, 16 Kibutz Galuyot Street, Netanya 42365, Israel. Email: atziner@netanya.ac.il

© International Association for Applied Psychology, 2002. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.
tended to demonstrate the superior reliability (Latham & Wexley, 1981) and validity (Latham, Fay, & Saari, 1979) of Behavioral Observation Scales (BOS) in comparison to Graphic Rating Scales (GRS). However, a comprehensive review of multiple rating formats incorporating both qualitative and quantitative criteria (Bernardin & Beatty, 1984) failed to find a single rating format that was uniformly superior in all respects—notwithstanding the conventionally accepted advantages of behavior-based scales (e.g. Fay & Latham, 1982; Feldman, 1981). Indeed, decades of research on the psychometric properties of various appraisal methods have not demonstrated conclusively, in terms of psychometric properties, the existence of a format that is most reliable or valid (cf. Gosselin, Werner & Halle, 1994; Landy & Farr, 1980; Latham & Wexley, 1981; Schwab, Heneman, & DeCotiis, 1975; Steiner, Rain, & Smalley, 1993; Tziner, 1990).

A complicating factor is that in most studies an accurate comparison of scale formats has not been possible because rating instruments have differed in terms of scale dimension content, scale anchors, and also in the number of scale points (cf. Borman, 1974). The inconsistency of past findings has also been attributed, in part, to the scarcity of studies that have subjected each type of scale to the same methodological scrutiny.

Tziner’s (1990) review does, though, permit more delimited distinctions to be drawn among the various formats. Evidently, forced choice scales are better able to minimise deliberate rating inflation, making them preferable for administrative purposes such as promotion, merit pay, and employment termination. Behavior-based formats, on the other hand, are generally superior to other formats in fostering performance improvement; when used with performance feedback, they tend to facilitate clarification of work roles for employees and the reduction of role ambiguity and conflict. Finally, with respect to psychometric characteristics, personnel comparison methods (such as paired comparisons and rank ordering) appear to be less susceptible to intentional leniency, central tendency, and range restriction biases.

**PERFORMANCE APPRAISAL FORMATS: EFFECTS ON RATERS AND RATEES AND WORKERS’ JOB PERFORMANCE**

Although it is difficult to recommend one appraisal format over another for multiple purposes, behaviorally-based formats do appear to offer advantages with respect to fostering performance improvement and worker development. Tziner and colleagues, via a 15-year, six-study research program, have investigated the effects of multiple methods and formats of performance appraisal on a variety of rater and ratee outcomes—measured subsequent to the evaluation process (Tziner, 1984; Tziner & Kopelman, 1988; Tziner & Latham, 1989; Tziner, Kopelman, & Joanis, 1997; Tziner, Joanis, &
Murphy, 2000). The primary goals of this research program were: (1) to throw light on the effectiveness and utility of various measuring tools; and (2) to assist human resource personnel, through the use of performance appraisal (PA), to contribute more effectively to the overall work environment and employee productivity in work organisations.

This research program was undertaken despite the scant research on the usefulness of PA methods for developmental and productivity-enhancing purposes. In broad terms, it has been found that PA, when administered properly, can serve as far more than an adjunct for administrative functions in areas such as remuneration or assignment decisions.

Past research has demonstrated that when performance feedback is precise and timely it may result in behavior change, even though job behaviors are generally difficult to modify (Campbell & Pritchard, 1976; Schneier, Shaw, & Beatty, 1991; Steers, 1975). Prior research has also indicated that performance appraisal can contribute to employee development (McGregor, 1957; Wexley, 1979) and the improvement of future job performance (Bernardin & Beatty, 1984; Murphy & Cleveland, 1991). Dorfman, Stephan, and Loveland (1986) demonstrated that performance appraisal can foster short-term behavior change through the following mechanisms: raters’ identification of employee strengths and weaknesses; the provision of performance feedback; and the facilitation of communication with supervisors. Yet, Tziner et al. (1997) noted that evidence regarding the long-term (future) effects of performance appraisal on employee attitudes and performance has yet to be established clearly (e.g. Nathan, Morhrman, & Milliman, 1991).

Murphy and Cleveland (1995) have also pointed to the relative dearth of investigations on subordinates’ reactions to appraisal systems as compared to the abundance of studies on the psychometric properties and accuracy of rating systems. Moreover, Bernardin and Beatty (1984) suggested that ratees’ reactions to an appraisal system might yield a more significant contribution to sustaining the viability of an appraisal system than its psychometric qualities. Notwithstanding the accuracy of the ratings or their freedom from errors, an appraisal system may become useless (and go unused) if it does not elicit positive reactions (e.g. satisfaction with the appraisal interview) among both raters and ratees (see also Hedge and Borman, 1995). More recent and impressive support for this position has been assembled by Cawley, Keeping, and Levy (1998) in a large meta-analytic study.

As described above, the two major types of PA rating tools currently used in the field are: (1) behavior-based rating formats which define job performance in terms of specific, “pinpointed”, observable actions (e.g. “ratee completes all sections of a crime report”), and (2) graphic rating scales (also referred to as GRS-based rating formats) that typically identify relatively broad and vague dimensions of performance in specific areas (e.g. “ratee pays increased attention to details”). In contrast to GRS scales which depend on
the subjective (and possibly biased) judgments of appraisers, behavior-based appraisals define job performance in relatively objective, quantitative terms. Clearly, the two most popular behavior-based appraisals are, respectively: (1) Behavioral Observation Scales (BOS), which ask raters to report the frequency of specific job-related behaviors, and (2) Behaviorally Anchored Rating Scales (BARS), which use behavioral statements to illustrate multiple performance levels. In particular, BOS clarifies for both the appraiser and ratee precisely what activities should (or should not) be performed on the job, in what manner, and how the outcomes link to performance ratings (Latham & Wexley, 1981).

Throughout the series of studies conducted by Tziner and colleagues, the researchers contended that these three particular rating scale formats (GRS, BOS, and BARS) might prove more or less useful in helping raters provide actionable feedback—viz., information that assists ratees in formulating plans and goals for acting on feedback. In light of the specific characteristics of these three scales (illustrated in the Appendix), it is possible to consolidate the supporting evidence to date. Thus, particular attention is given in the following account to the specific post-PA outcomes that served as dependent variables in Tziner et al.’s series of investigations.

The suppositions made by the researchers were as follows:

Communication

BOS appears more likely to minimise barriers in the communication process between superiors and subordinates because it pinpoints for both the appraiser and the appraisee specific organisational expectations and performance requirements. Consequently, role ambiguity and role conflicts are likely to be reduced and positive employee attitudes toward their work should be enhanced. Moreover, although merely indicative, it might be noted that Wiersma and Latham (1986) found that managers and employees preferred BOS (and the GRS) to Behaviorally Anchored Rating Scales (BARS).

Performance Goals

Because they differ in extent of focus on specific behaviors, and in the amount of information they potentially provide about actual behaviors, different rating scale formats are likely to affect the sorts of performance goals set by both raters and ratees. The GRS format is widely used by organisations for performance review purposes (Carroll & Schneier, 1982), yet such appraisals tend to be relatively vague and generic. Thus, the consequent courses of action derived from goals to be accomplished are also
likely to be “fuzzy” (lack goal clarity). However, it is reasonable to posit that rating scales that focus on specific behaviors (e.g. BOS) should lead to goals that are clearer, more directly observable, and better accepted by ratees than scales that provide less behavioral information. This enhanced clarity should in turn increase commitment to these goals (Carroll & Schneier, 1982; Dolan & Schuler, 1987; Huber, 1985; Pinder, 1984). Indeed, clarity, acceptance, and commitment are all recognised determinants of the success of performance appraisal and feedback (Latham & Wexley, 1977); consequently, these constructs were also employed as criteria in the Tziner et al. series of studies (see below).

Feedback

A behavioral observation scale (BOS-based) performance appraisal and review process should lead to higher levels of satisfaction and performance with the feedback process than would a graphic rating scale approach for three reasons. First, we recall, the BOS-based process conveys more precisely what an individual should do. As noted, rather than enjoining the individual to “improve your sense of responsibility”, and to “sharpen your leadership skills”, behavior-based feedback identifies the specific desired/undesired actions to be taken/avoided (Latham & Wexley, 1981). Second, BOS-based feedback is more likely to be accepted because it is seen as more factual, objective, and unbiased (Kopelman, 1986). Third, BOS-based feedback should be more conducive to setting performance goals that are specific and job-related, rather than vague and non-related, i.e. goals that minimise rater subjectivity (cf. Petit & Haines, 1994). These goal properties (specificity and relevance) should further influence the extent to which the appraisee perceives goals to be clear, acceptable, and worthy of commitment (Carroll & Schneier, 1982; Dolan & Schuler, 1987; Huber, 1985; Pinder, 1984).

User Satisfaction with the Appraisal Method

User satisfaction relates not only to feedback or goal setting; it concerns the overall appraisal process. The quality of this attitudinal concomitant is critical; if either the performance appraisal administrators or ratees are dissatisfied with the system, the effectiveness of the overall appraisal and feedback process is diminished (Gosselin & Murphy, 1994; Petit & Haines, 1994; Wiersma & Latham, 1986). Because behavior-based appraisals define job performance in relatively objective, quantitative terms, it should be seen by ratees as more accurate and fair, i.e. less prone to problems of rater bias, rater errors, and rater inexperience (Tharenou, 1995).
Rater Satisfaction

Rater dissatisfaction may emerge from anxiety related to the lack of desire or ability to rate one’s colleagues (among other causes). There is preliminary evidence that a BOS format helps decrease raters’ fear of possible confrontations with their subordinates (Latham, Fay, & Saari, 1979). This proposition is supported by the observation that raters who are generally uncomfortable carrying out an appraisal and delivering appropriate feedback (Arpin, 1994; Napier & Latham, 1986) may find that the use of BOS scales reduces this discomfort. This is likely to occur because a BOS-based performance appraisal and review (PAR) directs discussion toward the frequency of specific behaviors, rather than toward the rater’s evaluations of those behaviors. Moreover, insofar as ratees will likely view behavior-based ratings as relatively more factual and incontrovertible than GRS-based appraisals, raters should find the review process more congenial (Wiersma, Van Den Berg, & Latham, 1995). Indeed, to the extent that the performance review discussion is based on relatively objective and quantitative information, rather than ambiguous criteria and subjectively interpreted levels of performance, the resulting interaction should be less acrimonious and more satisfying to raters (and to ratees).

Raters are also likely to be more satisfied with BOS than with BARS because of the relatively objective nature of the BOS format which involves recording behavior frequencies, rather than extrapolating from actual to exemplar ratee behaviors (Latham & Wexley, 1977; Petit & Haines, 1994). Moreover, BOS seems easier both to understand and implement in an adequate manner (Wiersma & Latham, 1986; Latham & Wexley, 1981). Further, regarding the choice between BARS and BOS, Tziner (1990) noted that raters often have difficulty selecting the anchors that most accurately reflect a ratee’s performance; further, diametrically opposed (high vs low) statements can describe the same ratee’s behavior. In such cases, the rater may select an intermediate score as most representative, and concomitantly formulate a goal that is inappropriate. These drawbacks indicate that of the two behavior-based appraisal methods, the BOS format should result in more positive rater satisfaction reactions and higher ratee performance appraisal reviews (e.g. ratee satisfaction, clarification of path to goal) than appraisals based on the BARS format.

Tziner and colleagues conducted a series of investigations based on these premises, in which several of the drawbacks inherent in previous studies were overcome, and in which most of the above suppositions were tested empirically, some for the first time. The following review of these studies highlights some of the major findings and conclusions drawn from them, especially where they further elucidate theoretical conceptions or indicate avenues for future research in this area. By and large in these investigations, the three PA formats were compared with each other using samples of
supervisors (managers), trained to conduct performance rating, and ratees (subordinates) in real organisational settings. Except for the first investigation which focused on psychometric properties, the studies recorded the post-appraisal responses, reactions, and ratings of ratees (and in some cases of raters), according to the respective dependent variables being examined, as shown in Table 1.

The pertinent results and conclusions drawn by the researchers from each of these studies are summarised below. The “comments” represent the main ideas presented in the discussion of the individual papers.

**Tziner (1984a): Psychometric Properties of Rating Scales in a Real Organisational Setting**

*Description.* 57 managers in an Israeli aircraft plant appraised their subordinates using both BARS and a GRS format.

*Results.* In overall terms, the BARS method appeared less susceptible than the GRS to both halo and leniency effects; it also yielded marginally higher inter-rater agreement.

*Comment.* Results were due to the elimination of some methodological deficiencies inherent in former comparative studies (e.g. an equal number of scale points was ensured). This conclusion supported Bernardin and Smith’s (1981) observation that results not supporting the psychometric superiority of BARS may be accounted for by methodological flaws. This research did not, however, demonstrate that BARS is unequivocally superior to the GRS format regarding psychometric properties. For example, on some performance dimensions in this study, only marginal or non-significant leniency effect differences were found when comparing the two methods. (Yet, this finding may have been attributable to insufficient training of raters. Because rating errors are well-ingrained habits, evidence suggests that extensive rater training is necessary for avoiding such errors (Latham & Yukl, 1975). Indeed, Latham et al. developed a comprehensive workshop that: provided trainees with broad opportunities to practice the specified skills, provided trainees with feedback on their practice appraisal performances; and that ensured a comprehensive acquaintance with the appropriate behaviors to be observed.)

The results may also reflect contextual factors in the focal organisational setting. Some ratees evidently assumed that the appraisals were to be used for administrative purposes, which in turn might engender unwanted union pressures. That is to say, these kinds of contextual factors could have biased the ratings on the BARS, although they are generally regarded as more easily tenable compared to the GRS format. Despite the numerous
**TABLE 1**

Review of Dependent and Independent Variables in Tziner and Colleagues’ Studies on Effects of Performance Rating Formats on User Reactions (e.g. Rater Appraisal Satisfaction, Ratee Work Satisfaction) and on Psychometric Properties of Ratings

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>GRS vs</th>
<th>BARS</th>
<th>BOS</th>
<th>Additional independent variables</th>
<th>Hypotheses with respect to these dependent variables (outcomes):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tziner 1984a</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>Leniency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Halo effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Interrater R)</td>
</tr>
<tr>
<td>Tziner &amp; Kopelman 1988</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Performance feedback</td>
<td>Goal clarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal acceptance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal commitment</td>
</tr>
<tr>
<td>Tziner &amp; Latham 1989</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Performance feedback</td>
<td>Work satisfaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal-setting</td>
<td>Org. commitment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal clarity</td>
<td>Ratee appraisal—satisfaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal acceptance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal commitment</td>
<td></td>
</tr>
<tr>
<td>Tziner, Kopelman, &amp; Livneh 1993</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Goal-setting</td>
<td>Method performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal clarity</td>
<td>Ratee appraisal—satisfaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal acceptance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal commitment</td>
<td></td>
</tr>
<tr>
<td>Tziner, Kopelman, &amp; Joanis 1997</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Goal-setting</td>
<td>Rater appraisal—satisfaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal acceptance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal commitment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Path to goal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal commit’m’t</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal observab’ty</td>
<td></td>
</tr>
<tr>
<td>Tziner, Joanis, &amp; Murphy 2000</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Perception of goals</td>
<td>Ratee appraisal—satisfaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Goal observability</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Generally, it has been hypothesised that psychometric properties (e.g. leniency, halo effect) and user reactions (e.g. rater appraisal satisfaction, goal clarity, goal specificity, ratee organisational commitment) will be superior with BOS compared to BARS and GRS and of BARS compared with GRS.*
problems that inevitably arise in field research, such research, in contrast to laboratory studies (e.g. Fay & Latham, 1982), is essential for applying research on the results of training and/or type of rating scale effects on rating errors.

According to Feldman (1981), rating errors will probably creep into overall (global) evaluations of any type, independent of the rating method. Yet, the global performance evaluation, made after the ratings on performance dimensions were completed, was less contaminated (in Tziner, 1984a) by leniency effect in the BARS than in the graphic rating scale format. Tziner (1984a) noted, further, that due to the standardised frame of reference associated with observing, recording, and interpreting behavioral incidents (Bernardin & Smith, 1981), BARS-based ratings are less subject to the various rating errors described by Feldman (1981).

It appears, therefore, that the BARS method, as one of the broad cluster of behavioral approaches, prepares the rater better to summarise cognitively and abstract adequately from dimensional performance ratings, thus enabling their incorporation into a “well-designed” overall (global) evaluation. This quality should express itself in less susceptibility to various appraisal biases in performing an overall (global) evaluation. This proposition, though, that Tziner notes still has to be submitted to both theoretical elaboration and empirical examination. Future research should incorporate a measure of “true” performance, so that differential accuracy can be examined. Such a design would serve to replicate the primary findings of this study.

Tziner & Kopelman (1988): Effect of Rating Format on Goal-Setting Dimensions

Description. 10 managers and their 62 subordinates at an Israeli aircraft company participated in a field experiment that examined the effects of PA feedback derived from BOS and GRS scales. The criteria of interest were three goal-setting dimensions: goal clarity, goal acceptance, and goal commitment.

Results. The BOS format yielded significantly higher levels of goal clarity, goal acceptance, and goal commitment.

Comment. In a preliminary analysis of the data, an average intercorrelation of 0.60 was found among the three perceptual goal-setting dimensions. In light of the high interrelatedness among dependent variables, the three measures were analysed together using a multivariate analysis of variance (MANOVA). Results indicated significant differences in perceived goal dimensions across the two experimental groups, $F(1,60) = 51.14$, $P < 0.001$, Wilks’ lambda = 0.27.
A priori \( t \) tests (Wildt & Ahtola, 1978, p. 30) were used separately for the three goal-setting dimensions, the results of which indicated that ratees who received performance feedback using a BOS rating format reported significantly higher levels of goal clarity, goal acceptance, and goal commitment. This supports the notion that the behavior specificity of the BOS format results in the formulation of more focused goals that should, in turn, facilitate the improvement of ratees’ work behavior and job performance. In contrast, goals derived from the generic (and often vague) performance dimensions of GRS-based performance feedback, were less clear, acceptable, and yielded lower ratee commitment. Perhaps this “explains” why the performance appraisal process is often seen by supervisors as an unpleasant duty to be avoided, postponed, or—if inescapable—handled hurriedly (cf. Kopelman, 1986).

Tziner & Latham (1989): Effects of Appraisal Instrument, Feedback and Goal-Setting on Worker Satisfaction and Commitment

**Description.** In the light of studies showing that without trained appraisers it makes no difference what PA format is used (Fay & Latham, 1982), 20 managers received intensive training in appraising objectively, giving performance feedback, and setting specific goals when conducting performance appraisals. The sample consisted of 125 workers in the Israel Airports Authority in Israel. Following training, ratees were assigned randomly to one of four conditions: (i) feedback and goal-setting using BOS-based appraisals; (ii) feedback using BOS-based appraisals with no goal-setting; (iii) feedback and goal-setting using GRS-based appraisals; and (iv) feedback using GRS-based appraisals with no goal-setting. Work satisfaction and organisational commitment were measured before and after performance appraisals were conducted.

**Results.** Analysis revealed that: (a) BOS-based appraisals increased work satisfaction significantly more than did GRS-based appraisals; (b) feedback followed by goal-setting resulted in significantly higher worker satisfaction and organisational commitment than feedback alone, regardless of the appraisal scale that was used; and (c) the combination of BOS-based appraisal, feedback, and goal-setting led to significantly higher work satisfaction than in the other three experimental conditions.

**Comment.** The results of this field experiment reconfirm that performance feedback alone generates improvements to ratee organisational commitment and, particularly, to work satisfaction (e.g. Kim & Schuler,
1979). Importantly, PA consisting of performance feedback followed by goal-setting produced even more positive ratee outcomes than PA comprising only feedback.

Unlike work satisfaction, organisational commitment was only slightly affected by performance feedback alone. Organisational commitment has been defined as the extent to which an employee identifies with, and is involved in, his/her work unit and the organisation as a whole (Curry, Wakefield, Price, & Mueller, 1986). In view of the broader scope of organisational commitment in comparison to work satisfaction, it is not surprising that performance feedback had less effect. It would seem to follow that the determinants of organisational commitment are likely to be broader in scope than the results of performance feedback.

The process of goal-setting seems to give the appraisee a broader picture of the work unit and the organisation’s objectives, and how the set goals contribute to that unit and to the organisation’s goals. Consequently, feedback coupled with goal-setting was associated with higher organisational commitment.

As noted above, performance feedback combined with goal-setting contributed most strongly to ratees’ measures of work satisfaction, possibly since goal-setting fosters feelings of participation in work-related issues (Miller & Monge, 1986; Schneider, 1981) and meaningfulness at work, both of which were components of the study’s satisfaction measure.

Regarding the comparison of method formats, the BOS-based appraisal was superior in comparison to the GRS-based appraisal in every instance. The indigenous specificity of BOS (in comparison to GRS) was cited as a factor contributing to higher work satisfaction. The specificity likely strengthened ratees’ feelings of control at work and dissipated ambiguity about expectations and requirements.

It might be noted that this study as well as Tziner and Kopelman’s (1988) investigation extend research on Behavioral Observation Scales to the Israeli organizational setting which differs largely from its North American counterpart in terms of work value structure (Falbe, Nobel-Ben Yoav, & Tziner, 1986). The positive reactions to BOS relative to alternative instruments such as GRS found in both laboratory and field settings in the US (Bernardin & Beatty, 1984; Wiersma & Latham, 1986) have now been replicated in Israel.

Tziner et al. suggested further development and testing of a variety of appraisal instruments that are relevant and free of contamination, in a number of different organizations and jobs. This empirical process would provide raters too cynical to risk antagonizing their subordinates with negative feedback (Latham, 1986; Napier & Latham, 1986; Wiersma & Latham, 1986) with reliable instruments, and would further the corroboration, clarification, or extension of the underlying conceptual frameworks and
mechanisms of the PA process. Moreover, Tziner et al. recommend the extension of the study beyond the effects of appraisal format and performance review on work-related attitudes to the impact of these variables on employee behavior such as absenteeism, turnover, and job performance.


Description. In a natural extension of the previous studies, 16 nurse-managers used BOS and GRS scales for a performance appraisal and review (PAR) process with 115 nurse-subordinates in Israel. Levels of goal clarity, acceptance, and commitment were compared, as were levels of appraisal process satisfaction and improvement in job performance.

Results. It was hypothesised that the BOS-based PAR process, in comparison to the GRS-based PAR, would yield higher levels of: (1) goal clarity, acceptance, and commitment; (2) satisfaction with the appraisal process; and (3) subsequent job performance. All three predictions were confirmed.

Comment. The BOS-based process of performance appraisal and review again produced consistently higher levels of goal clarity, acceptance, and commitment compared with the GRS-based approach, and higher levels of satisfaction with the appraisal process. The researchers noted that the latter finding was compatible with Mount’s (1983) observation 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 the formulation, performance, and facilitation of personal development plans.

With respect to the effect on rated performance, mean performance scores were essentially the same in the GRS and BOS conditions at “Time 1” (5.11 and 5.10, respectively). Regarding subsequent changes in rated performance, none was found using the GRS condition at “Time 2” (5.13, \( r(106) = 0.71 \)), whereas overall performance in the BOS condition improved modestly, to 5.22, \( t(104) = 1.12, P = 0.13 \). Examination of the changes in individual performance, however, indicated a significant improvement in the BOS condition, to \( t(52) = 2.77; P < 0.01 \), and only an insignificant improvement in the GRS condition, \( r(53) = 0.71 \).

Tziner et al. noted that given the widespread use of GRS-based performance appraisals (Teel, 1980), these results give credence to Nathan et al.’s (1991) observation that the typical PA review may not change performance.
Yet, in this study, the BOS-based approach did lead to significant (albeit modest) improved job performance (on a within-person basis).

It was suggested that future research obtain data amenable to confirmatory factor analysis, and, further, that data on relatively objective goal properties such as specificity and job-relatedness be obtained using reliable external assessment of goal statements. Such information, it was argued, would facilitate the process of confirming the overall causal model of linkages between goal properties, perceived goal characteristics, process satisfaction, and improved job performance. These advances would go a long way in clarifying the long ignored processes triggered by the two rating formats, BOS and GRS (Murphy & Cleveland, 1991). 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.

Tziner, Kopelman, & Joanis (1997): Investigation of Raters’ and Ratees’ Reactions to Three Methods of PA

Description. In a further extension of the previous studies, attention was given to all three appraisal formats and to the reactions of raters, as well as to ratees. It was hypothesised that, in comparison to a GRS-based process, BARS- and BOS-based PARs would generate (i) superior performance goals (in terms of their characteristics of observability and specificity), (ii) superior ratee perceptions of set goals (clearer clarification of path to goal, more acceptable, more amenable to commitment/achievement), and (iii) more favorable user reactions (ratee and rater satisfaction with the PAR process). Additionally, because of the limitations inherent in BARS, it was posited that both rater and ratee satisfaction would be higher using BOS compared to BARS. It should be noted that in this study, as in the previous field investigations, raters were trained extensively in all aspects of the PA and feedback processes prior to administering the forms.

“Specificity” of goals was assessed in terms of the extent to which precise courses of action were spelled out (explicit/concrete and desirable/undesirable behaviors/objectives were pinpointed), and “observability” in terms of the extent to which the goal described a behavior or action that could be detected by an external observer. These goal properties were judged independently of raters and ratees, and the concordance between three judges, using the Kappa coefficient, was $K = 0.67$ for specificity and $K = 0.72$ for observability. The ratings of the three judges were averaged to calculate separate scores for specificity and observability for each of the goals.
Results. The following results were found:

Goal Specificity. BOS-based reviews yielded significantly higher goal specificity than that of either the GRS or BARS. The latter finding was anticipated, in view of the limitations of BARS described above.

Goal Observability. None of the three review formats demonstrated an advantage with respect to goal observability, perhaps because the observability of a goal largely reflects the nature of the work situation (Riggio & Cole, 1992), which is not affected by the appraisal format.

Path to Goal. The BOS-based review did not yield higher clarification of path to goal than did GRS; however, the difference was in the predicted direction and approached statistical significance ($P = 0.063$).

Goal Acceptance and Goal Commitment. The BOS-based review yielded higher levels of goal acceptance and goal commitment than did both the GRS and BARS methods.

Ratee Satisfaction. BOS-based reviews yielded higher ratee satisfaction than those of GRS and BARS.

Rater Satisfaction. There was no support for the proposition that behavior-based reviews would yield higher rater satisfaction than GRS, possibly because this variable is so reflective of the overall rating context. Given that the ratings were explicitly conducted solely for developmental purposes and—it was emphasised—were to have no administrative implications, this might have obviated the tensions that often accompany a GRS-based review.

Comment. With respect to user reactions, these findings clearly reinforce the prior research findings concerning the superiority of a BOS-based performance appraisal review over a review that is either GRS- or BARS-based.

Concerning future research, Tziner et al. recommended the continued collection of data from various sources (as was the case in this study where goal properties were assessed by independent judges, performance review satisfaction data were provided by raters, and goal characteristics and attitudinal data were provided by the ratees). Future research should extend this line of inquiry to include additional cultural settings beyond Canada (this study), Israel (Tziner & Kopelman, 1988), and the Netherlands (Wiersma et al., 1995), and to match data from all sources, thus permitting causal modeling of the entire nomological network.
Tziner, Joannis, & Murphy (2000): A Comparison of Three Methods of Performance Appraisal with Regard to Goal Properties, Goal Perceptions and Ratee Satisfaction

Description and Results. In a similar comparison of GRS with BOS and BARS rating formats, performance improvement goals for police officers using BOS were judged by experts to be the most observable and specific. In addition, ratees’ satisfaction with performance appraisal was highest, and their perceptions of performance goals most favorable, when using BOS. The overall results suggest that behaviorally oriented rating formats can enhance the developmental applications of performance appraisal. With respect to specific outcomes, the results were as follows:

Specificity. As predicted, the goals set with BOS were more specific than those established by the use of either BARS or GRS. As previously noted, the BOS format asks the rater to record the frequency with which specific critical behaviors occur, increasing the likelihood that goals will also be framed in terms of specific, observable behaviors. Both BARS and GRS formats ask raters to provide general evaluations. (With BARS, behavioral statements are used merely to illustrate what scales and scale levels mean.) Thus with BARS and GRS formats it is likely that ratee goals will be tailored toward affecting raters’ evaluations rather than toward carrying out the specific behaviors listed on the BOS.

Observability. In terms of goal observability, however, no reliable difference was observed between BOS and GRS, although both of these methods were superior to BARS.

Ratee Perceptions. Ratee perceptions of goals were more favorable with BOS and GRS than with BARS, and no significant differences were discovered between GRS and BOS with regard to ratee perception of goals. Although BOS and GRS appeared to be equivalent in terms of ratee perceptions, it could be argued that the increased clarity of goals obtained when using BOS may be a sufficiently important advantage to lead one to prefer BOS to GRS methods.

Ratee Satisfaction. The level of ratee satisfaction with the appraisal system itself was also higher for BOS and GRS than for BARS. No significant differences were discovered between GRS and BOS with regard to ratee satisfaction with the appraisal system. Wiersma and Latham (1986) also found these two methods to be as good as each other by virtue of their being equally easy to use; they also showed that individuals were more satisfied with BOS than BARS because it can minimise personality disputes, assist...
in the explanation of weak assessments, make feedback easier, and render appraisal sessions more comprehensible.

Comment. In this particular study, Tziner et al. used a simple measure of ratee perceptions of goals, which involved the aggregation of the disparate dimensions of goal clarity, goal acceptance, and goal commitment. The literature points, however, to certain ambiguities and contradictions in the various definitions of these three terms, some authors asserting that acceptance and commitment are interchangeable concepts (e.g. Huber, 1985), while other researchers conclude that the two notions have different meanings (Locke, 1968).

Locke has defined goal acceptance as a state of mind where individuals agree to attain the goals that have been set, while commitment refers to the degree of resolve to reach these goals. The person setting these goals is not discussed, but the a priori acceptance on the part of the ratee is implied. In a later development, Locke, Shaw, Saari, and Latham (1981) defined commitment more broadly, applying it to any type of goal, independent of its source. Thus, the goal may have been set by the individuals themselves, their superiors, or in collaboration with others. Acceptance is subsumed under commitment, and is seen as referring specifically to the degree of commitment to assigned goals; while commitment here implies goal acceptance (Locke, Latham, & Erez, 1988). This description of the conceptual relationship was borne out by Early and Kanfer (1985).

All told, there still remain persistent ambiguities in the definitions of the variables labeled goal acceptance, goal commitment, and goal clarity. It would be beneficial to undertake a more extensive study in order to determine the extent to which these constructs are distinct, conceptually and empirically.

Another acknowledged limitation of goal setting stems from the accent placed upon individual results, a practice that often engenders a mentality of “every man/woman for himself/herself”, and the minimisation of cooperation for attaining collective goals. Teamwork, which is important in so many settings (such as the police, Kane & Freeman, 1987), thus becomes problematic, an observation borne out by Petit and Haines (1994) who noted the difficulty of applying goal-setting procedures to several job types.

Finally, Tziner et al. suggest that it would be interesting to employ a cost–benefit analysis approach to the use of these methods. In effect, one is obliged to ask how BOS can be recommended when GRS is easier to use and less costly to develop, and yet seems to generate similar results—at least with regard to ratee satisfaction. Tziner et al. concluded with a core question for human resource personnel: Are the satisfaction of the ratees and their reactions to the goal set for them enough to justify the time, energy, and effort required for the implementation of BOS? (It should be noted, however, that

© International Association for Applied Psychology, 2002.
Tziner et al. did not consider performance/individual development benefits which are of considerable importance to organisations today.)

GENERAL CONCLUSIONS

From an examination of the above comments that reflect the findings of the individual studies, several generalised conclusions may be drawn regarding the effects of different rating scales.

1. Although in the overall picture, behavioral rating methods appear to have the advantage over graphic scales, largely due to their precise nature, the specific drawbacks of BARS indicate that it is less preferable than either of the other two formats. This contention was borne out by all the studies in which BARS was always held in the lowest favor and produced the least impressive contributions to the improvement of work habits in the workplace.

2. The psychometric advantages of BOS over GRS seem to be superior reliability and validity. Because of its specificity (and its clearer format), however, BOS also appears to minimise communication barriers, role ambiguities, lead to clearer and more observable goals and consequent ratee commitment to carry them through. The feedback process with BOS is generally more focused, factual, objective, and unbiased, and thus more likely to garner specific job-related goals that positively influence appraisee attitudes to the overall PA process, and their consequent work satisfaction. Of importance was the finding that goal-setting adds incrementally to feedback, especially with the use of BOS, although the advantage of BOS over GRS was not substantiated in terms of organisational commitment.

   Raters are also more likely to feel comfortable with BOS because it directs attention to the frequency of specific and incontrovertible behaviors, in contrast to ambiguous and relatively subjective criteria of GRS formats.

3. Regarding goal properties, Tziner et al. have argued for obtaining relatively objective data concerning goals, such as specificity and job-relatedness. These goal properties can be obtained using reliable external assessments of goal statements. Moreover, in the light of the conceptual confusion over two facets of perceived goal characteristics (i.e. acceptance and commitment), further research should seek to determine whether these variables are actually part of the same concept, or whether it is possible to differentiate between them.

4. Concerning goal-setting, it was suggested that in addition to setting individual goals, supervisors should attend to goals that reflect group-level performance.
5. Despite the overall impression of the superiority of BOS over GRS, several of these investigations indicated little difference between BOS and GRS formats with respect to particular dependent variables. These were: “goal observability” (Tziner, Kopelman, & Joanis, 1997; Tziner, Joannis, & Murphy, 2000), “path to goal” (Tziner, Kopelman, & Joanis, 1997), and “rater satisfaction” (Tziner, Kopelman, & Joanis, 1997; Tziner, Joannis, & Murphy, 2000), and ratee perception of goals (Tziner, Joanis, and Murphy, 2000).

In the light of the above, and the relative ease and lower cost of producing GRS formats, it can be argued that before the above variables (outcomes) are examined, and before opting for one or other appraisal method, there may be a case for adopting a cost–benefit analysis approach. As mentioned, ratee satisfaction may not provide sufficient justification for the time, energy, and effort required implementing BOS.

6. Of particular importance to organisations, especially in light of accelerating global competition, is the impact of scale format on average levels of job performance. Only one study examined this effect (Tziner, Kopelman, & Livneh, 1993). Although average levels of performance were not significantly higher using the BOS format, the mean difference approached statistical significance. However, examined on a within-person basis, the effect on individual performance was a significant improvement using BOS, \( P < 0.01 \) versus essentially no change in the GRS condition. It is this finding that, if replicated, would perhaps justify the added expense of developing BOS-based performance appraisal instruments.

7. The inferences made about raters unprepared to administer PA because of the potential negative consequences to themselves emphasise the underlying operational aspect of this research. Independent of the academic discussion of superiority of rating formats, there is the very real consideration of the development of reliable and valid appraisal instruments in a number of organisational and cultural settings, the use of which could eventually make or break an organisation. Data should also be obtained from multiple sources.

8. One of the interesting issues to emerge from these studies concerns the relationships among the various components of the theorised nomological network. More specifically, data should be obtained that permit examination of the overall causal model of linkages between goal properties, perceived goal characteristics, process satisfaction, and improved job performance. Repeatedly, Tziner et al. assert that future research should be based on data amenable to confirmatory factor analysis (LISREL) that would go a long way in clarifying “the long ignored processes” triggered by the two rating formats, BOS and GRS.
Presently, however, it is impossible to perform LISREL analyses due to the absence of necessary data.

9. Finally, notwithstanding the decades of research on the psychometric properties of rating formats, Tziner and colleagues have argued that the real payoff of their line of research which is relatively distinct may well be in terms of changes in on-the-job work behavior and job performance. Moreover, they contend that the research would do well if extended into the other areas of work behavior and job performance, such as absenteeism and turnover.

REFERENCES


1 Note: For this reason they have reviewed extensively and intensively their work in this article.

© International Association for Applied Psychology, 2002.


© International Association for Applied Psychology, 2002.


APPENDIX: EXAMPLES OF GRS, BOS, AND BARS FORMATS

Behavioral Observation Scale Format (BOS)

<table>
<thead>
<tr>
<th>Professional knowledge</th>
<th>Updates the pilot with flight information well ahead of time</th>
<th>1 2 3 4 5 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Navigates planes through clouds relying on radar guidance</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>Tolerance</td>
<td>Despite ‘nagging’ pilot, he proceeds to reason his decisions to him</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td>With arrival of a ‘strange’ plane (private plane or a pilot landing for the first time), he meticulously explains regular landing instructions</td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>Performance under stress</td>
<td>When pilot reports an emergency, he instructs him with great precision while directing air traffic as usual</td>
<td>1 2 3 4 5 6</td>
</tr>
</tbody>
</table>

6 = to a very great extent or almost always
1 = to a very little extent or almost never

or

6 = 91%–100% of time the ratee exhibited the behavior
5 = 71%–90% of time the ratee exhibited the behavior
4 = 51%–70% of time the ratee exhibited the behavior
3 = 31%–50% of time the ratee exhibited the behavior
2 = 11%–30% of time the ratee exhibited the behavior
1 = 0%–10% of time the ratee exhibited the behavior
### Graphic Rating Scale Format (GRS)

#### A. Level of performance

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Extremely Good</th>
<th>Very Good</th>
<th>Good</th>
<th>Fairly Good</th>
<th>Fair</th>
<th>Fairly Low</th>
<th>Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

#### B. Ability to adjust to work

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Extremely Good</th>
<th>Very Good</th>
<th>Good</th>
<th>Fairly Good</th>
<th>Fair</th>
<th>Fairly Low</th>
<th>Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
### Behaviorally Anchored Rating Scale Format (BARS)

<table>
<thead>
<tr>
<th>Scale point</th>
<th>Performance dimension</th>
<th>Quality of work performance</th>
<th>Organisational commitment</th>
<th>Professional knowledge</th>
<th>Quantity of work</th>
<th>Initiative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>The quality of work is below the average of others in similar positions in his unit</td>
<td>Only slightly interested in his work, and is inclined at times to neglect it</td>
<td>The standard of his professional knowledge falls below that required for his position and therefore needs supplementing and improving</td>
<td>Amount of work performance in a pre-specified time interval, is lower than that of others in the same or similar positions</td>
<td>Only occasionally suggests new ideas for improvement of work processes and assignments on his own initiative</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>The quality of his work is usually on a par with those in similar positions in his unit</td>
<td>Can be relied on to carry out his assignments. The level of his involvement and identification with the job is on a par with others in similar positions in his unit</td>
<td>The standard of his professional knowledge is adequate for his position</td>
<td>Amount of work performed in a pre-specified time interval is equal to that of others in the same or similar positions</td>
<td>Usually suggests new ideas for improvement of work processes and assignments on his own initiative, although he does not make an effort to follow through their application</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>The quality of work is usually above the standard of those in similar positions in his unit</td>
<td>Interested in his work and aims at improvement. The level of his involvement and identification with the job surpasses that of others in similar positions in his unit</td>
<td>His professional knowledge in his job is comprehensive and of a high standard</td>
<td>Amount of work performed in a pre-specified time interval is occasionally higher than that of others in the same or similar positions</td>
<td>Usually suggests new ideas for improvement of work processes and assignments on his own initiative, although he occasionally has difficulty in carrying them out</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>The quality of his work is outstanding; it excels in accuracy, order, and integrity</td>
<td>Shows above average willingness to forgo personal benefits for the sake of his job, showing full identification and involvement with his work</td>
<td>His professional knowledge in his job is comprehensive, up-to-date, and of an exceptionally high standard</td>
<td>Amount of work performed in a pre-specified time interval is usually higher than that of others in the same or similar positions</td>
<td>Usually suggests new ideas for improvement of work processes and assignments on his own initiative</td>
</tr>
</tbody>
</table>