Performance Appraisal: Methods and Rating Errors

Fred C. Lunenburg
Sam Houston State University

ABSTRACT

Performance appraisal is the systematic observation and evaluation of employees' performance. Some of the most commonly used performance appraisal methods include the judgmental approach, the absolute standards approach, and the results-oriented approach. Ideally, performance appraisal should be completely accurate and objective. However, the performance appraisal process is far from accurate and objective, sometimes resulting in rating errors. Common rating errors include strictness or leniency, central tendency, halo effect, and recency of events.

Virtually every organization has a formal employee performance appraisal system. *Performance appraisal* may be defined as the process by which superiors evaluate the performance of subordinates, typically on an annual or semiannual basis for the purpose of determining raises, promotions, or training needs (Grote, 2011).

There are a number of alternative performance appraisal methods, each with their own strengths and weaknesses that make them more appropriate for use in some situations than in others (Dessler, 2012). Further, it is assumed that superiors accurately appraise their subordinates' performance, leading to unbiased and objective judgments. However, the performance appraisal process is far from objective, sometimes resulting in rating errors (Guerra-Lopez, 2009).

Performance Appraisal Methods

Organizations currently use several methods to appraise performance. For the sake of simplicity, we can group them into three categories: the judgmental approach, the absolute standards approach, and the results-oriented approach (Bratton, 2012; Dessler, 2012).

Judgmental Approach

A manager or performance appraiser is asked to compare an employee with other employees and rate the person on a number of traits or behavioral dimensions. These appraisal systems are based on the exercise of judgment by the superior. Four widely used judgmental approaches are graphic rating scales, ranking, paired comparison, and forced distribution.

Graphic rating scales. A popular, simple technique for evaluating employees is to use a *graphic rating scale*. Table 1 shows a typical rating scale for a manager. Note that the scale lists a number of important work dimensions (such as leadership and management) and a performance range for each one. For each work dimension, the evaluation scheme is typically used to assess the important work dimensions: (1) unacceptable, (2) needs improvement, (3) acceptable, (4) commendable, and (5) outstanding. The assigned values for each dimension are then added up and totaled.

Table 1

Abbreviated Graphic Rating Scale for Managers

Work Dimension	Unacceptable	Needs Improvement	Acceptable	Commendable	Outstanding
Leadership	1	2	3	4	5
Management	1	2	3	4	5
Personnel administration	1	2	3	4	5
Administrative teaming	1	2	3	4	5
Budgeting	1	2	3	4	5

Ranking. An alternative method to graphic rating scales involves managers ranking their subordinates in order of their performance effectiveness from best to worst. The usual procedure requires the rater to write the name of the best subordinate on the top of a list, then the name of the worst at the bottom and continue this sequential procedure until all subordinates are listed. *Ranking* is most frequently used for making decisions such as promotions or the merit salary increase each employee will receive.

Paired comparison. A modification of the ranking procedure is the *paired comparison* technique. The method overcomes the problem associated with differentiating between subordinates in the middle range of the distribution. Under paired comparisons, raters compare only two subordinates at a time until all two-way comparisons have been made among all employees. After rating all pairs, the manager can put the subordinates into a rank order by counting up the number of times each employee has been judged superior.

Forced distribution. "Grading on a curve" is a good example of the forced distribution method of performance appraisal. With this technique, the rater places a predetermined percentage of rates into four or five performance categories. For example, if a five-point scale is used, the manager might decide to distribute employees as follows: 5 percent in the "unacceptable" category, 25 percent in the "needs improvement" category, 40 percent in the "acceptable" category, 25 percent in the "commendable" category, and 5 percent in the "outstanding" category. The usual procedure for accomplishing such a distribution is to record each employee's name on a separate index card. Then, for each dimension being appraised (leadership, management, etc.), the employee's index card is placed in one of the five categories.

Absolute Standards Approach

Most appraisal measures that employ an absolute standards approach are based on job analysis. As discussed earlier, this type of analysis can provide a more detailed description of the actual behavior necessary for effective performance. Managers compare the performance of each employee to a certain standard instead of to the performance of other employees; thus, they rate the degree to which performance meets the standard. The most common performance appraisal processes in this group are checklists, essays, critical incidents, and behaviorally anchored rating scales.

Checklists. The most common technique in the absolute standards group is some sort of *checklist*. Checklists tend to be more behaviorally based than either graphic rating scales or other employee-comparison methods. Table 2 presents a humorous example of a checklist that might be used to appraise managers' performance. More elaborate procedures, such as weighted and force choice checklists, are also available. Specific weights are assigned to a list of work behaviors in the *weighted checklist*. A *forced choice checklist* consists of job-behavior statements with two to five response items in each set that correlate with high- and low-performing employees. The end result is a single numerical rating that is useful for personnel decisions such as salary and promotion (Rothwell, 2012).

Table 2

A Guide to Appraising Managers' Performance

Performance Factor	Outstanding	High Satisfactory	Satisfactory	Low Satisfactory	Unsatisfactory
Quality	Leaps tall buildings with a single bound	Needs running start to jump tall buildings	Can only leap small buildings	Crashes into buildings	Cannot recognize buildings
Timeliness	Is faster than a speeding bullet	Only as fast as a speeding bullet	Somewhat slower than a bullet	Can only shoot bullets	Wounds self with bullets
Initiative	Is stronger than a locomotive	Is stronger than a bull elephant	Is stronger than a bull	Shoots the bull	Smells like a bull
Adaptability	Walks on water consistently	Walks on water in emergencies	Washes with water	Drinks water	Passes water in emergencies
Communication	Talks with God	Talks with angels	Talks to himself	Argues with himself	Loses those arguments
Relationship	Belongs in general management	Belongs in executive ranks	Belongs in rank and file	Belongs behind a broom	Belongs with competitor
Planning	Too bright to worry	Worries about future	Worries about present	Worries about past	Too dumb to worry

Essays. The *essay* method requires the rater to describe in writing each employee's strengths and weaknesses, along with suggestions for ways to improve performance. Some organizations require every rater to respond to specific open-ended questions, whereas others allow more flexibility. Compared to employee comparison methods, the essay method is time-consuming and difficult to quantify. Variations in the writing skills of raters are another limitation. Some organizations have combined the graphic and essay methods by providing space for comments on the graphic rating scale.

Critical incidents. The *critical incidents* technique begins by identifying job requirements for successful performance. Job requirements are those behaviors that determine whether the job is being done effectively or ineffectively. The manager keeps a log, for each subordinate, of both effective and ineffective "incidents" of on-the-job

behaviors. The incidents are then analyzed and refined into a composite picture of the required essentials in a particular job. From this a checklist is developed, which constitutes the framework against which the subordinate is evaluated. During the evaluation conference, the manager can refer to the critical incidents to correct work deficiencies, identify training needs, or praise successful performance.

Behaviorally anchored rating scales. A newer and somewhat related approach to the critical incidents technique is the *behaviorally anchored rating scale* (BARS). It was developed to cope with the problem of identifying scale anchor points. Specifically, the scale points such as unacceptable, needs improvement, acceptable, commendable, and outstanding (as shown in Table 1) may be difficult to define and may lead to unreliable or invalid appraisal results (Deblieux, 2003). Hence, the BARS define scale points with specific behavior statements that describe varying degrees of performance. The form for a BARS generally covers six to eight specifically defined performance dimensions. A BARS should be developed for each dimension.

Figure 1 shows an example of a BARS for the testing competence-performance dimension for industrial psychologists. The scale anchors define the particular response categories for the evaluator. The response made by the evaluator is specific enough to be used as feedback in an appraisal interview with the industrial psychologists and is meaningful to the subordinate (Fletcher, 2009). For example, if the industrial psychologist were given a 3 on this dimension, the subordinate would be given the specific performance indicators that led to the evaluator's rating.

Job: Industrial Psychologist
Performance Dimension: Testing

Outstanding Performance	5	This industrial psychologist is recognized as an expert and can be expected to help others and to provide counsel to others working on the team.
Commendable Performance	4	This industrial psychologist can be expected to know almost everything about testing and can provide assistance in solving difficult problems.
Acceptable Performance	3	This industrial psychologist can be expected to work diligently on normal caseload and to complete them on time.
Needs Improvement	2	This industrial psychologist can be expected to work late on testing to keep up with the caseload.
Unacceptable Performance	1	This industrial psychologist is confused and can be expected to hinder the completion of the caseload because of lack of knowledge.

Figure 1. Behaviorally anchored rating scale.

Results-Oriented Approaches

In recent years, results-oriented approaches to performance appraisal have been suggested as an alternative to the judgmental and absolute standards approaches. As the name implies, the emphasis of results-oriented approaches is on the evaluation of results—both quantitative and qualitative. Put another way, the focus is on what the subordinate is supposed to accomplish on the job rather than a consideration of the subordinate's traits or on-the-job behaviors (Fletcher, 2009; Grote, 2011).

Goal Setting. One popular results-oriented approach is *goal setting*. Goal-setting is often associated with motivation and, more specifically, as a motivational technique (Locke & Latham, 1994). Goal setting can also serve as the foundation for an organization's performance appraisal system. It is particularly well suited to high-level management positions for which methods such as BARS may be inappropriate.

This program typically includes two major elements. First, the supervisor and the subordinate meet to discuss goals, which are established by the supervisor alone or jointly by the supervisor and the subordinate. Second, the supervisor and the subordinate meet to appraise the subordinate's performance in relation to the previously established goals. For example, suppose a manager sets a goal of increasing product sales by 15 percent. At the end of the fiscal year, this goal provides a framework for performance appraisal. If sales have increased by 15 percent or more, a positive performance appraisal is likely. However, if sales have increased by only 5 percent and if the manager is directly responsible for the results, a more negative evaluation may be in order. Then suggestions for improvement can be specified.

Other results-oriented measures. Besides goal setting, managers can use a variety of other results-oriented measures to assess subordinate performance. Some suggestions include measures of quantity of output, such as number of sales, words typed, or items produced; measures of quality, such as reputation of the product, typographical errors, or items rejected; measures of lost time, such as absenteeism or tardiness; or measures involving education, training, or experience, such as time in the field or time in a particular position. Although these measures tend to be nonjudgmental, they measure only one dimension of job performance. Such measures can also be tied to a goal-setting program.

Rating Errors

In conducting performance appraisals, managers must be careful to avoid making rating errors. Four of the more common rating errors are strictness or leniency, central tendency, halo effect, and recency of events (Deblieux, 2003; Rothwell, 2012).

Strictness or Leniency

Some supervisors tend to rate all their subordinates consistently low or high. These are referred to as *strictness and leniency errors*. The strict rater gives ratings lower than the subordinate deserves. This strictness error penalizes superior subordinates. The lenient rater tends to giver higher ratings than the subordinate deserves. Just as the strictness error punishes exceptional subordinates, so does the leniency error. Strictness-leniency bias presents less of a problem when absolute standards and results-oriented approaches to performance appraisal are used.

Central Tendency

Some raters are reluctant to rate subordinates as very high or very low. They dislike being too strict with anyone by giving them an extremely low rating, and they may believe that no one ever deserves to get the highest possible rating. The result of this type of attitude is that everyone is rated around average. Figure 2 depicts examples of strictness, leniency, and central tendency biases. The distribution of ratings on the left of the figure indicates a strictness error; those in the middle indicate a central tendency error; and the cluster on the right indicates a leniency error.
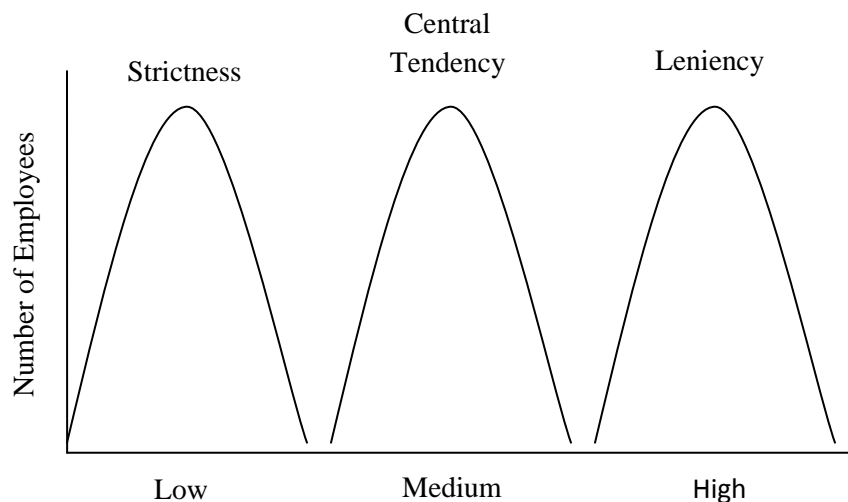

Figure 2. Strictness, central tendency, and leniency performance ratings.

Halo Effect

When a single positive or negative dimension of a subordinate's performance is allowed to influence the supervisor's rating of that subordinate on other dimensions, a *halo effect* is operating. For example, the supervisor likes Tom because he is so cooperative. The halo effect leads Tom's supervisor to automatically rate him high on all appraisal dimensions, including leadership, management, personnel administration, administrative teaming, and even budgeting. The result is that subordinates are rated consistently high, medium, or low on all performance appraisal dimensions.

Recency of Events

Ideally, performance appraisals should be based on data collected about a subordinate's performance over an entire evaluation period (usually six months to a year). However, as is often the case, the supervisor is likely to consider recent performance more strongly than performance behaviors that occurred earlier. This is

called the *recency of events error*. Failure to include all performance behaviors in the performance appraisal of a subordinate can bias the ratings.

Strictness or leniency, central tendency, halo effect, and recency of events all result in inaccurate performance appraisals of employees. The absolute standards and results-oriented approaches to performance appraisal, particularly BARS and goal setting, attempt to minimize such rating errors.

Conclusion

Performance appraisal is the systematic observation and evaluation of employees' performance. Some of the most commonly used performance appraisal methods include the judgmental approach, the absolute standards approach, and the results-oriented approach. Ideally, performance appraisal should be completely accurate and objective. However, the performance appraisal process is far from accurate and objective, sometimes resulting in rating errors. Common rating errors include strictness or leniency, central tendency, halo effect, and recency of events.

References

Bratton, J. (2012). *Human resource management: Theory and practice*. New York, NY: Palgrave Macmillan.

Deblieux, M. (2003). *Performance appraisal source book: A collection of practical samples*. Alpharetta, GA: Society for Human Resource Management.

Dessler, G. (2012). *Human resource management*. Upper Saddle River, NJ: Prentice Hall.

Fletcher, C. (2009). *Appraisal, feedback, and development: Making performance review work*. New York, NY: Routledge.

Grote, R. C. (2011). *How to be good at performance appraisals: Simple, effective, done right*. Cambridge, MA: Harvard Business Review Press.

Guerra-Lopez, I. J. (2009). *Performance evaluation: Proven approaches to improving program and organizational performance*. New York, NY: Wiley.

Locke, E. A., & Latham, G. P. (1994). *A theory of goal setting and task performance* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Rothwell, W. J. (2012). *Encyclopedia of human resource management set-3 volumes*. New York, NY: Wiley.