Chapter 6

Rating Error Traps

I don't know that there are any shortcuts to doing a good job.

—Sandra Day O'Connor, first female justice on the U.S. Supreme Court

When employees trust their supervisors to conduct fair and unbiased evaluations, their satisfaction with the system increases dramatically.[1] But, the process must remain trustworthy; if it breaks down toward the end because of rating errors, the whole effort is tainted. Both employees and supervisors must be aware of what these errors are and how they affect performance evaluations. Employees need to stay vigilant, and managers' good intentions must be accompanied by the skill, understanding, and training required to reduce rating errors as much as possible.

Even then, it's difficult. Although software can support some types of evaluations, performance appraisals are a human process--conceived, developed, and administrated by people. No evaluation comes with a flaw-free guarantee. Peter Drucker makes it clear that performance cannot be measured fully. "As each human being is unique, we cannot simply add them together, or subtract them from one another. … [T]o arrive at meaningful measurements is one of the greatest challenges to management. "[2]

Goal-based systems are often seen as the best current option for rating performance, but they must be used carefully. The kinds of behaviors that are specified in the goal-setting process are exactly what the employee will tend to focus on in his or her work, so it's critical that these are the behaviors the organization wants to encourage.[3]

Examples can be drawn from the composite employees of the three scenarios used at the ends of chapters in this book. Because Marilyn takes great pride in being cited for outstanding interpersonal relationships, there is some danger that she may not insist that some work move as fast as required. Richard's goals of processing a high number of duplicating orders per week may undercut quality and, in fact, may hurt business. Although Peg's poor performance must be addressed at every level, it's possible that paying attention to effective telephone protocols is undercut by a responsibility such as supporting typing overloads. This is another instance in which customer service may suffer.

Rater Bias

Focus on the dynamics of worker characteristics appears to have surfaced close to a century ago. Workplace awareness of individual needs and differences goes back to at least the early 20th century. In England, the work of Charles Darwin popularized ideas that individuals differed from each other in ways that were important. In France, the work of Alfred Binet and Theophile Simon led to the development of the first intelligence tests, and during World War I several armies tried using these tests to better assign soldiers to jobs. By 1923, Personnel Management was spelling out how to match a person's skills and aptitudes with job requirements.[4]

Performance evaluations have been evolving since the early 20th century, when they were created, in part, to reduce the potential for labor unrest. Companies began introducing formal job analysis to aid in employee selection and rationalize the hodgepodge of wage rates that existed in many companies.[5]

Bias in the process is well documented. A 15-year study by Pennsylvania State University revealed a pattern of employment practices that historically helped men get promoted to upper-management positions. It showed that employers make decisions based on “impression management,” the ability of employees to shape and manage a self-image that positively influences others. Women were shown to be “low self-monitors,” that is, less concerned with crafting an impressionable image than men, whose “chameleon-like quality” helped them adapt to changing social climates. “When employee promotions are based on subjective evaluations rather than skills and talent, men have the edge, with a 15 percent higher chance of being promoted.”[6]

Another study, at a Midwestern university, found that when viewing a female employee on video, students consistently gave her lower ratings on such characteristics as “dependability” and “ability to do the job” when she was pregnant than they did when she was shown five months later, even though her behavior was the same on both tapes..[7]

Evaluation problems also emerge because of perceptual differences in definitions. When words such as poor, fair, adequate, satisfactory, and excellent are used, the evaluation can be distorted. Exactly what does each mean? In comparison with whom? Is every employee being rated by the same standard?

Common Rating Errors

Halo effect. An error can occur when an outstanding quality becomes the basis of an entire rating—but only when the rating is not justified. There is a difference between halo errors and a true halo, which is justified by across-the-board excellent performance. Some organizations ask raters to evaluate everyone on a single dimension before proceeding to the next. The aim is to encourage raters to focus on a particular dimension rather than overall performance. Another often-used method is “reverse wording,” which structures forms so that a favorable answer for the first question might be ten on a scale of one to ten, while a favorable response to the sixth question might be one on the same scale. Again, the evaluator is required to focus on each question separately. Favorable first impressions that stay intact despite evolving problems are sometimes attributed to the halo effect**.**

Horns effect. In this reversal of the halo effect, a negative dimension becomes the basis of the whole evaluation, and a poor rating emerges because the negative performance in one area brings down all the others. If a sales manager, for example, receives a poor rating because he or she turns in paperwork late, that rating might be extended to sales skills that are excellent. Careful documentation of sales closings through the year would be one way to correct that kind of unjustified evaluation.

Sunflower effect. Managers may worry that giving employees a rating of “average” will reflect poorly on them, and give all their employees top ratings. But that can backfire: The managers’ supervisors may question the ratings and conclude that the managers did not spend enough time on the review to do it carefully. During an exit interview at one association, the departing employee, when asked his perspective on performance appraisals, responded that they were a joke—despite his earning among the highest ratings in the entire association. “Don’t get me wrong,” he said. “I like the increases, but I’m not doing my best work. I want to learn more, and either my supervisor doesn't care about understanding the work of different employees, or she just doesn't care."

Leniency or harshness error. One rater may tend to be more lenient or tougher with employees than other managers; several raters may all have different value systems. When appraisals apply words like "adequate" and "good," standards might not be defined clearly enough to ensure consistency throughout an organization. Some raters just tend to mark high; others low. It's often why employees report feeling as if they're back in school—and frustrated. Two employees performing similarly may receive quite different ratings from their respective supervisors simply because of these supervisors' tendencies to rate high or low. This error is also called positive or negative leniency. Evaluators report that positive leniency "motivates employees and makes them feel good."[8] They may justify negative leniency with the observation that "nobody's perfect." The tendency to make leniency errors is particularly strong when raters are rushed.[9] Raters may be asked to check their evaluations for a pattern toward leniency. One common remedy for this is to ask raters to distribute their ratings, with percentages designated for the number of employees rated as excellent, good, adequate, and poor. But this is a contrived solution and can unjustly place employees in false categories simply because their evaluators must respond to percentage demands.

Central tendency error. Some supervisors are reluctant to give high or low ratings. They rate all employees as average and fail to distinguish between the star performers and those who need specific support. Also called the clustered ratings error or scale shrinking, this method absolves managers from having to make judgments. It's sometimes used by raters who feel they don't know an employee well enough to come up with an actual rating. Sticking to the middle makes these evaluations less useful for making personnel decisions such as promotions, salary increases, training, counseling, and even feedback. Raters who make a central tendency error can be shown the bigger picture so they understand how their ratings are distorting the evaluation process. Sometimes organizations ask that employees be ranked so they don't all end up in the middle. But imposing a bell curve disbursement of employees' ratings can create other problems. Forcing the hand of a manager in an effort to arrive at predetermined ratings, or at a distribution that supports the increased budget, is not fair to employees and can lead to serious morale and legal problems.

Sugar-coating error. Discussing concerns verbally isn't enough to evaluate performance well. Problems develop when supervisors talk at length about needs for improvement and other concerns, but just jot down a few general lines on the appraisal form itself. Everything communicated verbally should also appear in writing, and vice versa. If that is not done, and further action needs to be taken, the available documentation falls short. For example, an employee may consistently give other departments incorrect information, but her appraisal reads only, "more care is needed in communicating to others." That does not cover the full scope of concern.

Recency of events error. Alleviating this error is one excellent reason for ensuring ongoing documentation and discussion. Without these activities, raters can forget the last five months of behavior and evaluate just the past five weeks. Employees sometimes exploit this reality, becoming especially active and visible just before review time. But even if a supervisor documents performance effectively, employees will benefit from carrying a "picture " of their efforts during the review cycle into the appraisal meeting, particularly if he or she has shown improvement since the previous cycle.

Critical incidents effect. Similar to the halo and horns effects, this error distorts the overall review by giving undue emphasis to a single episode, positive or negative. No one incident should dominate the entire review cycle. An especially excellent or poor performance at any point in the cycle should not subsume performance during the rest of the cycle.

Contrast effect. When the evaluation of one employee affects that of another, it's known as a contrast error. Because every employee merits an appraisal based on individual performance, the contrast error skews the process. If a stellar performer, for example, is evaluated right before a good performer, the contrast might demote the second employee to just a fair rating. A contrast error can also result when the rater compares past and present performance. An employee rated "good" in one review might be rated "poor" in the next one, even though his or her performance would otherwise be rated fair.

Personal Bias Error. Bias has many faces -- and none belongs in the appraisal. Some types of bias are readily apparent. But others are subtler, such as what's called a similarity error. This distortion occurs when a supervisor gives a higher rating to an employee simply because the two of them share similar characteristics, such as getting to work on time or being willing to work late. A supervisor may be totally unaware that he or she is even doing this. Long-time employees should be evaluated based on the quality of their performance, not on the number of years they have been with an organization. Ongoing, careful documentation, coupled with objective standards, goes a long way toward countering a bias error. Stereotyping, or generalizing across a group, is tied to bias. Just like other forms of bias, stereotypical views can also be subtle. Not recognizing individual differences and assuming, for example, that all marketing directors are motivated and ambitious and that every engineer is highly analytical creates subjective standards that can lead to rating errors. There may also be perceptual difference errors shaped by the manager's experiences and perspectives. A teacher, for example, might be appraised based on the rater's classroom experiences rather than on his or her own teaching skills.

Low motivation error. Evidence shows that it is more difficult to obtain accurate appraisals when important rewards depend on the results.[10] When the stakes are high, supervisors may be reluctant to provide an unbiased appraisal for fear of hurting the employee's chances of receiving added compensation, a promotion, or other opportunity for professional growth.

Past anchoring errors. Employees get caught in this error when managers rate performances based on prior evaluations instead of taking a fresh look.

Sampling error. This error occurs when the evaluator rates an entire review cycle on the basis of just a small sample of an employee's work.

Varying standards error. When two or more employees perform similar work yet are held to different standards, the discrepancy distorts a fair and just evaluation process. One employee, for example, might be rated "good " for closing 65 percent of his or her sales while another employee documenting the same number of closings is rated only "fair. "

Holding employees accountable when it's not their fault. Widespread negative evaluations probably mean that fault is with management, which is not holding all employees accountable. If, for example, documents from large numbers of staff tend to show up late at certain times, there may be problems with technology, not the employees. It's important to check that, before appraisal time. Employees should also not be held accountable for work requirements they were never told about, a problem that can surface when work standards are set without referring to the job description and actual requirements. This error is tied closely to the setting unrealistic objectives, such as holding employees accountable for work they're not trained for, or having them field multiple priorities with unrealistic deadlines, complete complex assignments without adequate research assistance, and similar misplaced goals.

Attribution bias. Distorted ratings occur when outstanding performance is "attributed" to factors external to the employee being rated, such as "great team support," but poor performance is perceived as the result of an employee's own behavior. Poor technological acumen, for example, might be attributed to lack of employee understanding, while excellent technological skills might be due to organizational training. In one instance, the employee is held accountable; in the second, an external factor gains the praise. A supervisor may want to grab credit for good performance. In other words, when work goes well, good management is credited. When it doesn't, poor employee performance is blamed.

Reducing Rating Errors

Management Review

Performance appraisals should be reviewed by the manager’s manager or the Human Resources department, before the manager sits down with the employee. This helps ensure that the rating matches the narrative and that the overall assessment is justified. Being one step removed from the manager/employee relationship, the manager’s manager can contribute the objective insight and constructive feedback needed to make necessary changes. The management review provides safeguards that are important to both the supervisor and the organization.

Sidebar

Checks & Balances

- ✓ Conduct rater training.
- ✓ Clarify the sequence of evaluation procedures.
- ✓ Ensure that the supervisor’s supervisor signs off on all appraisals before they’re given to employees.
- ✓ Build in time for the Human Resources department to conduct a second check.
- ✓ Provide a clear form inviting the employee to respond, in writing, to the review.
- ✓ Institute a formal appeals review, including a mediator if the employee desires one.

Training

"Research shows that training can minimize rating errors. When raters learned which data to focus on, how to interpret it, and how to use it to formulate judgments, ratings were more reliable and accurate than when there was no training or training incongruent with rating needs," wrote E.D. Pulakos.[11] The two most popular types of training programs are designed to help eliminate the kinds of errors described above and to improve supervisors' observation and recording skills. While programs dealing with errors seem to eliminate many of them from ratings, there is much less evidence that this kind of training actually increases the accuracy of appraisals. Programs focused on observation and recording skills may offer greater improvements in accuracy than those that simply focus on errors, according to an article in *The Journal of Applied Psychology*.[12]

Regardless of the format, effective training conveys a full understanding of the evaluation process. It involves learning how to complete all materials, becoming a true player in appraisal sessions, being alert to legal implications, and leaving the door open to a continuing process. Trainers design sessions reflecting the particular culture and needs of organizations, with workshop activities that tackle the tough challenge of reducing rating errors.

Rater training serves to remind supervisors and employees of the importance of the appraisal process. It serves to underscore that for every employee to receive the fairest rating possible, information must be reviewed responsibly, knowledgeably, and legally. Because it's not difficult for rating errors to creep into any performance evaluation, it's essential to be on guard against any distortion of the process.

Training will work only if supervisors are motivated and committed to applying it and organizational accountability is built into the process. Conducting high-quality, timely performance appraisals should, in fact, be a rated factor in the appraisals of supervisors.

Appeals Process

Most companies build employee response into the evaluation process. This may take the form of an employee-generated memo, but often it is an official response form that is attached to the written review. Sometimes happy employees will respond accordingly, but employees who are displeased with their ratings are more apt to express their reaction. Some organizations make sure those employees have another option: a complaint resolution process or formal grievance procedure. Smart companies believe that any employee with a complaint should have the opportunity to be heard and have the concern promptly and objectively reviewed and corrected, if necessary, with no fear of retaliation. Organizations support resolution through varying routes; some are two- or three-step processes, others are more extensive.

For any process to be effective, it should include an orderly and well-communicated system, specific steps and time frames that are followed by any employee with a similar issue. Conflict resolution processes begin with an employee putting his or her rebuttal in writing, then discussing those concerns with the immediate supervisor. The organization should designate in writing how many days the employee has to schedule and conduct this meeting and each subsequent meeting as necessary.

If the issue is not resolved by this point, a more formal process begins. An independent mediator might be brought into the process to facilitate a resolution as quickly as possible. If the employee remains dissatisfied, the supervisor can schedule a private meeting between the employee and the supervisor's supervisor. This meeting, too, should occur within a specified time, after which both managers write the results of their respective meetings on designated portions of the appraisal form. If an employee still seeks a resolution, he or she can meet with the next level of management. That individual looks into the matter and proposes a solution.

If the dispute is still unsettled, a panel of trained peers, guided by a mediator, can explore the information from both parties. The peers will make a decision within organizational parameters. Human Resources and one of the top executives will review the panel's decision, either signing off on the decision and making it binding, or offering an alternative solution that all parties can live with.

Three Scenarios

The following evaluation incidents about Marilyn, Richard, and Peg demonstrate how easily rating errors can occur. Because Marilyn, a generally high performer, gets along well with everyone, her supervisor gives her an across-the-board "excellent" rating, without first reviewing her performance across the full range of rating categories **(halo effect).** As another example, one of Marilyn's key responsibilities during the appraisal cycle was to contract out construction work designed to expand the office in time for new attorneys to come on board. Despite contractor delays, Marilyn made sure the job was finished on deadline. As significant as it was, her supervisor allowed this one achievement to color all others **(critical incidents effect).**

Richard's parent was rushed to the hospital two weeks before his performance review. Unexpectedly away from the office, he never delegated his work, and part of an important assignment slipped through the cracks. This incident stayed fresh in the mind of his supervisor, who gave him an overall low performance rating despite Richard's other positive achievements during the appraisal cycle **(recency of events error).** In another instance, Richard's duplicating department budget was sliced, despite a commitment to a client to produce work requiring technology that now had to be placed on hold. Told to deliver the work anyway, Richard had to outsource the work, then was held responsible for the budget deficit **(unrealistic objectives error).**

Despite Peg's continuing poor performance, her supervisor keeps trying to support her, postponing or completely derailing the need to fire her. She's also concerned that a poor rating will affect her supervisor's review of her as a manager. So she gives Peg a higher rating than her performance deserves, a disservice to Peg, her supervisor, and the advertising agency that employs them **(sunflower effect).**

Sidebar

Exercises for Large & Small Groups

Open by asking participants to generate a list of why performance appraisals are important; this jump-starts thinking about the review's long-term importance *before* focusing on the form.

Ask what documentation is needed to arrive at a fair rating, including last year's goals, significant accomplishments, job description, work examples, observation notes, input from others, employee’s self-assessment, and any other job-or goal-related information based on performance that should not be overlooked. Build a discussion about the value and pitfalls of each item.

Stage a role-play in which an “employee” works through each section of the evaluation form, with ratings based on examples drawn from the documentation. The “employee” challenges the ratings, providing the “supervisor” a chance to clarify why and how he or she arrived at the ratings. Sprinkled through the explanation are blatant, then subtler examples of rating errors. Participants have sheets listing many rating errors. Individually, then as a group, they are identified.

Small-group exercises are useful, too. After a general explanation of rating errors, written scenarios are distributed and participants break into small groups to identify the source of problems. The larger group can then convene to discuss reactions to the process of identifying, then correcting errors. The group can discuss the best ways for supervisors and employees to spot rating errors before they take hold.

Footnotes

1. Shimon L. Dolan and Denis Moran, "The Effects of Rater-Ratee Relationship on Ratee Perceptions of the Appraisal Process," *Human Resource Management*, 8th ed., (New York: McGraw-Hill/Irwin, 2001), 337-351.
2. Drucker, 81-82.
3. DeNisi and Griffin, 250.
4. DeNisi, 7.
5. Sanford M. Jacoby, A Century of Human Resource Management, from a paper presented at the 75th Anniversary Conference of Industrial Relations Counselors, Princeton University, September 11, 2001.
6. David Day et al., *HR Briefing*, (Aspen Publishers, July 1, 2002), 7.
7. Jane Halpert, Midge Wilson, Julia Hickman, "Pregnancy as a Source of Bias in Performance Appraisals," *Journal of Organizational Behavior* 14 (1993), 655.
8. William S. Swan, *How To Do a Superior Performance Appraisal*, (New York: John Wiley & Sons, 1991), 120.
9. Ibid., 121.
10. H. John Bernardin and Richard W. Beatty, *Performance Appraisal: Assessing Human Behavior at Work,* (Boston, MA: Kent Publishing, 1984), 140.
11. E.D. Pulakos, “The Development of Training Programs to Increase Accuracy on Different Rating Forms,” *Organizational Behavior and Human Decision Processes*, 38 (1986), 76-91.
12. Jerry W. Hedge and Michael J. Kavanagh, "Improving the Accuracy of Performance Evaluations: Comparisons of Three Methods of Performance Appraiser Training," *Journal of*

Applied Psychology, 68-73, in *Human Resource Management* (8th ed.), John M. Ivancevich, McGraw-Hill/Irwin, NY, 2001, 268.