IN THIS ARTICLE
Performance Appraisals, Upward Appraisals, 360-Degree Feedback

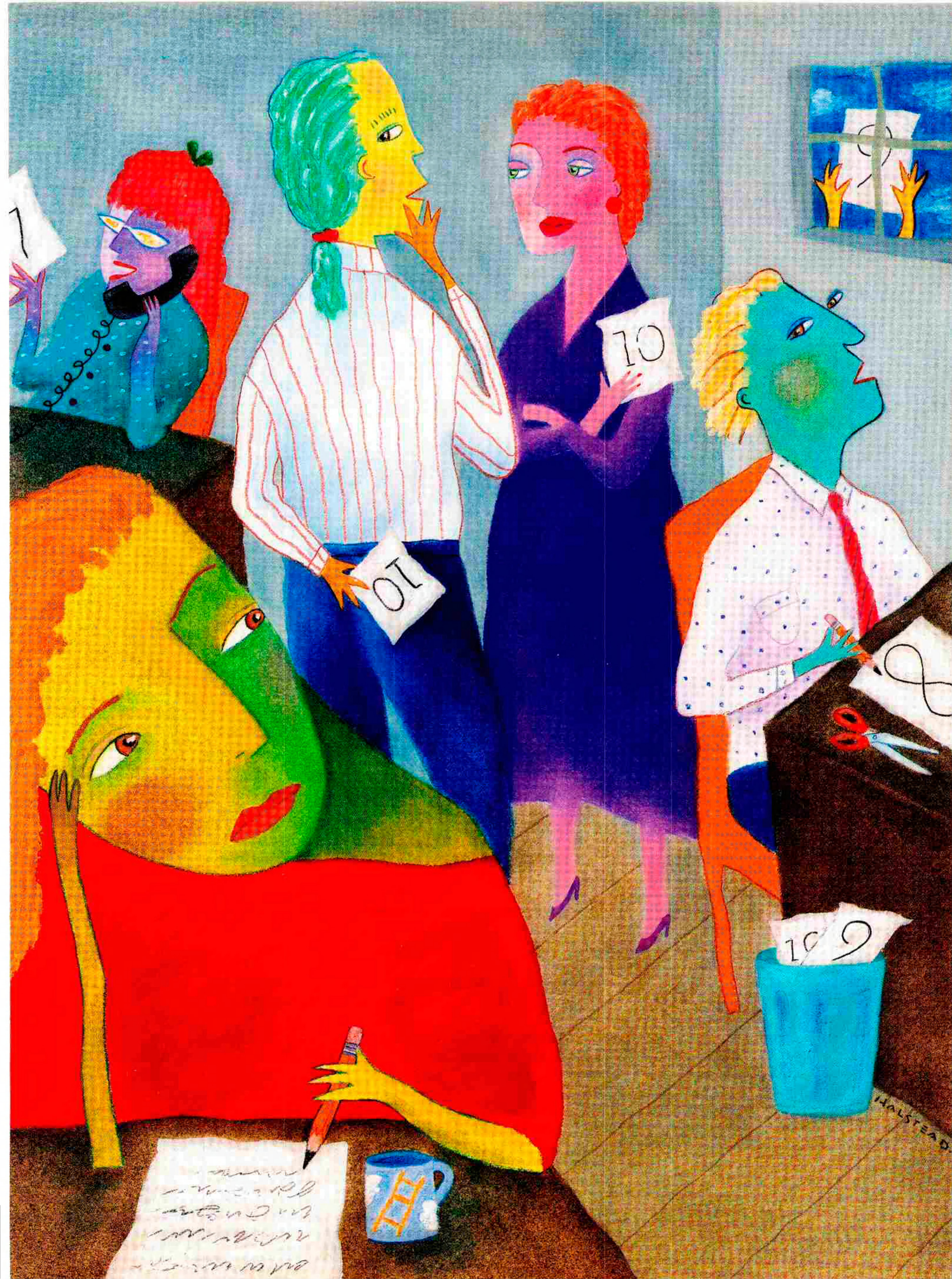
What factors contribute to the success or failure of multirater feedback systems?

Straight Talk About Multirater Feedback

By David W. Bracken

They go by many names: "Multirater systems," "multi-perspective ratings," "upward feedback," and "full-circle feedback" are just a few. Whatever you call them, multirater systems are apparently on the rise, as more and more organizations design and implement processes in which employees are rated by some combination of managers, supervisors, peers, direct reports, and even customers. • In some organizations, the raters' input is used solely for developmental purposes. But the trend is moving rapidly toward the use of multirater systems for performance appraisal, as well as for other personnel actions, such as succession planning and outplacement. • The trend appears to be more than a fad. The multirater process is taking hold in many major organizations as an integral part of their management processes. Raters and ratees know that the feedback will be used for performance appraisal and other personnel decisions that are important to most employees. They also know that the systems place accountability for action and follow-through squarely on the shoulders of the person being rated. So employees are likely to take multirater systems seriously. • The potential bad news in all of this is that we know very little about multirater processes. In particular, we don't know much about the ability of raters to provide reliable information about the performance of their co-workers, supervisors, or suppliers.

Illustration by Virginia Halstead


10
HALSTEAD

Some managers may choose to dwell on the legal ramifications of the situation. But a compelling argument for learning more about multirater systems before implementing one is that these systems affect people's lives. That's true whether the information is used for performance appraisal, succession planning, outplacement, or "just" development.

In turn, the decisions we base on multirater feedback ultimately improve or impair organization effectiveness, because they influence the quality of personnel decisions, organization change efforts, and employee commitment.

What factors contribute to the success or failure of multirater systems? To answer that question, let's first look at the definition of a successful multirater system:

- It is reliable, providing consistent ratings.
- It is valid, because it provides feedback that is job related.
- It is easy to use, understandable, and relevant.
- It creates positive change at both the individual level and the organizational level.

The multirater process

We can learn more about multirater systems from several sources. We can glean useful information from the substantial body of research on effective appraisal systems, survey methods, and rater characteristics. The rest of our understanding must come from our personal experience and research on multirater systems.

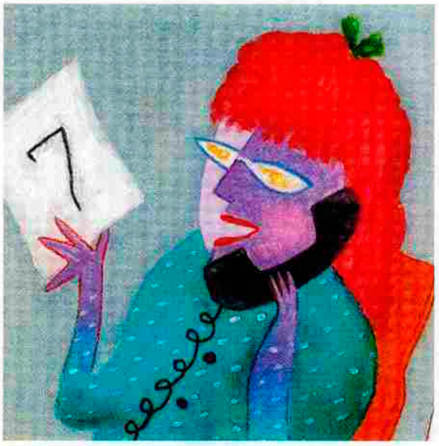
Opportunities to succeed or fail in implementing a multirater system occur at every stage of the process:

- process design and planning
- instrument development
- instrument design
- administration
- feedback processing and reporting
- action planning as a result of feedback.

Each of those phases is reviewed below, with a focus on the kinds of errors that can invalidate ratings, reduce confidence in the system, and result in lost opportunities for creating change.

In addition, system planners should pay attention to real or perceived opportunities for unfairness to be created in the system. Unfairness occurs when practices vary across the organization during implementation of the multirater process or in the use of information generated by the process. In other words, it's important to determine whether everybody is playing by the same rules.

For this discussion, assume that the target population is a supervisor or manager who will be rated by direct reports and customers (including internal customers). But many of the points also apply to other employee groups who could be rated by people inside or outside the organization.



DETERMINE WHETHER EVERYBODY IS PLAYING BY THE SAME RULES

Process design and planning

Three issues are key in the first stage of putting a multirater system into practice:

- Who will the raters be?
- When should you collect their input?
- How do you maintain confidentiality throughout the process, so that raters and ratees can feel assured of process integrity?

First, let's discuss the identification of raters. Who should provide the feedback? Proponents of true full-circle rating systems advocate the inclusion of raters from every possible perspective, including some raters from outside the organization, such as customers. The potential problems with that approach are both logistical and philosophical.

Organizations that try to implement true 360-degree processes soon find themselves awash in forms (if paper is used). Raters may have to spend an unreasonable amount of time on the evaluations, regardless of the technology used. (Some solutions to that problem are discussed later in this article).

One potential pitfall of allowing a manager to select a lot of raters is the increased possibility of putting evaluation forms into the hands of people who have a limited opportunity to observe that manager. Some managers who have been allowed to distribute unlimited numbers of forms have chosen to collect input from 30 or more raters! Numbers like that have to raise questions about the quality of those ratings.

The whole point of full-circle rating systems is to gather information from raters with varying perspectives on the performance of the manager. But there is legitimate concern about the quality of ratings from some groups, particularly external customers.

External customers may not have the same motivations as employees to provide high-quality input. A more practical problem is the difficulty of structuring questions to ask external customers. The questions must be written so that the responses of external customers can be combined with those of internal raters. But external customers will have little information on such topics as the leadership and performance-management skills of the ratee. More work needs to be done to understand better the role and characteristics of external-customer feedback in multirater systems.

Another issue to consider in the process-planning stage is timing. An effective multirater system will be administered on a schedule that meets two criteria:

- Collection of feedback should be timed so that the results can be incorporated into appraisals and other planning systems, if appropriate.
- Ratings should be gathered when most employees will have an opportunity to participate; for example,

you shouldn't try to collect feedback during major vacation periods or holidays.

The perception that confidentiality has been maintained throughout the process is critical to the success of a multirater system. It influences rater's response rates and honesty in providing feedback. And it affects a targeted manager's commitment to acting on the results of the ratings.

The most direct way to establish confidentiality is through the use of an outside person to process the information. Then, you should reinforce the promise of confidentiality by a public commitment from senior management that completed forms will not be read by any internal personnel. (Instead, raters' input will be compiled in aggregate form before it is seen by anyone at the organization.)

Instrument development

The most effective personnel systems—including selection methods, appraisal processes, and succession plans—tend to be those that reflect the vision and values of the organization of which they are a part. And an organization's vision and values reflect the place where the organization wants to be, not necessarily the place where it is today. That may differ from the traditional definition for "job relevance," which has historically described work as it is currently performed.

A multirater system should reflect both the present and the future in describing the behaviors expected of an organization's management team.

The questions on the feedback instrument should focus on behavior, not general traits. In other words, the questionnaire should ask raters to report whether the manager does or does not do something (a behavior) rather than whether the manager possesses some personal characteristic. Discerning a trait is more difficult and subjective than identifying a behavior; it entails the observation of multiple behaviors across a variety of settings.

The behaviors included on the instrument can be developed with the participation of employees—both management and nonmanagement—through facilitated group discussions.

Those behaviors should flow directly from the organization's vision and values, and the survey should present them in that fashion. In other words, survey designers should group behaviors under category titles that correspond to the organization's vision or value statements. This serves to "operationalize" the vision and values, giving life to abstract concepts such as "respect for the individual" and "customer first."

See the box, "Sample Questionnaire Items for Rating a Supervisor," for examples.

An important part of a questionnaire item is the response scale (assuming that the instrument is in the multiple-choice format). My research has shown that response scales do make a difference in the ways that respondents answer various questions, and that the popular Frequency scales (which include such choices as "always," "usually," "sometimes," and "never") have some serious deficiencies that should preclude their use.

Satisfaction scales (with choices ranging from "very satisfied" to "very dissatisfied") and Agreement scales ("strongly agree" to "strongly disagree") yield data that tend to be more helpful.

Offering about six choices for responses creates enough variance so that it is possible to detect behavior changes over time, as managers begin to incorporate their ratings into their day-to-day work lives. And it's best to offer an even number of alternatives. This prevents raters from "waffling" in their opinions, as people are likely to do when a midpoint choice is available.

At times, some raters really won't be able to answer with any of the available options. For those situations, it's better to provide a separate, well-identified "don't know" or "not applicable" choice, rather than asking respondents to leave the question blank.

Instrument design

Developers and administrators of multirater systems should make every effort to encourage full participation and complete responses from every employee who is asked to be a rater.

Response rates are particularly critical for managers with small work groups, who may not receive a report at all unless all of the employees in a work group return completed surveys.

In any size work group, partially completed questionnaires can prevent managers from receiving the benefit of feedback on the full skill set addressed by the questionnaire items. Depending on how the instrument is organized (in categories or at random), scores may be affected or unavailable for entire categories of behaviors if respondents leave questions blank.

Various elements of questionnaire design can affect response rates. Chief among them is length. When questionnaires are too long, more employees tend to leave them only

---

Sample Questionnaire Items for Rating a Supervisor

Questions might take the following form on a multirater-assessment instrument that has employees rate their supervisors. For each item, the rater must assign a rating to his or her manager's behavior.

On the sample survey items below, choices include "very satisfied," "satisfied," "dissatisfied," and "very dissatisfied."

A section with the category title "Respect for the Individual" might include the following question:

How satisfied am I that—

- My manager is open to different opinions.
- My manager supports a balance between work and personal life.
- My manager provides me with opportunities to grow and develop on the job.

Under the category "Customer First," the survey might ask the following question:

How satisfied am I that—

- My manager has a good understanding of our customers' needs.
- My manager recognizes excellence in customer service.
- My manager responds quickly to customer questions and complaints.

partially completed—or will not return them at all.

Length can be even more critical if your process involves some employees in filling out questionnaires on more than one ratee. This is especially likely if your system calls for ratings by peers as well as subordinates. Don't expect employees to spend more than 10 to 15 minutes filling out a questionnaire. That usually translates to an instrument of 40 to 60 items.

Open-ended questions can be useful for the respondent who wants to qualify or expand on some multiple-choice responses. And the data such questions provide can be helpful for the person being rated, who must first interpret the results and then act on them.

One useful format for open-ended questions asks respondents to complete phrases such as these:

- My manager should stop doing...
- My manager should start doing...
- My manager should continue doing...

When compiling the ratings, have such comments typed verbatim, and include them in the manager's report. Some organizations will do some editing of "essay" questions, but that can be risky. If employees perceive "censorship" of their responses, the credibility of the process will be harmed.

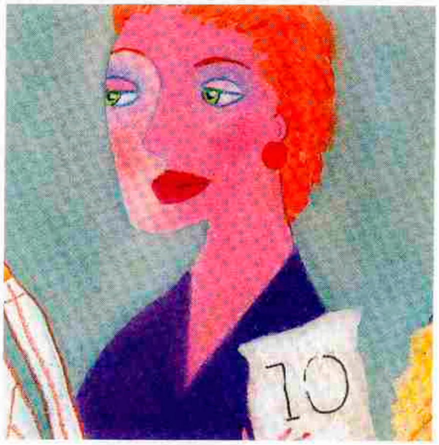
Process administration

Every employee should feel encouraged to participate in the multirater process. Often, personnel systems of various kinds leave out certain groups of employees, including night-shift workers and employees in decentralized locations. Employees receive an entirely different type of signal if top management does not participate in the process.

A multirater process should apply across the organization—horizontally, vertically, and over time. Any perception—or reality—that the organization has denied some group of employees equal opportunity to participate can hinder the system in three ways:

- It can result in managers receiving incomplete information in their reports.
- It can create the perception that results may be incomplete, inaccurate, and biased.
- It can create dissent among segments of the employee population, and reduce their levels of support for the process.

Organizations can encourage greater participation by providing time during working hours for completing the questionnaires. The best way is to hold group sessions at which managers or process facilitators can give directions and answer questions. If employees will be mailing in their completed surveys, be sure to provide postage so that employees can return their surveys



PAPER-AND-PENCIL IS BY FAR THE MOST POPULAR SURVEY METHOD

directly to the person or firm who will process the information.

An effective upward-feedback process (one in which subordinates rate managers) generally requires that all managers participate as ratees, and that all employees have the opportunity to participate as raters of their managers. Organizations that use peer ratings as well face a more complex challenge: getting representative feedback without overwhelming the organization with huge numbers of forms distributed and huge amounts of time spent filling them out.

The opportunity for a ratee to "stack the deck" by choosing only "friendly" raters is another potential problem with a system that allows a ratee to choose the people who will receive questionnaires. To avoid this pitfall, the organization should follow these three tips:

- Specify both a minimum and a maximum number of forms that ratees can distribute to their peers.
- Require every ratee to have another party review his or her distribution list. This other party could be, for example, a manager one level up, or someone in human resources.
- Allow individual employees to opt out of the rating process after they have received a certain number of forms to complete. (At least one organization has automated the process, in order to prevent employees from receiving too many questionnaires to complete.)

Document and publish all such procedures and policies regarding your rating system. If people feel that the rules governing the composition of rating groups vary across the organization, they are likely to think (rightly or not) that the process is unfair. Actually, some research indicates that it is actually quite difficult to stack the deck. But perception is what's important here. And employee perceptions can hurt the process if planners don't take care during the implementation phase.

A good multirater process will include some sort of rater training, especially the first time the process is administered. If the organization administers the questionnaires to groups of employees on company time (which is the recommended method), you can train the raters before they receive the instruments. The content of the training:

- an overview of the purpose of the process and the rationale behind it
- a discussion of how designers came up with the instrument
- details on how the collected information will be used
- information on administration methods and time lines
- assurances about how confidentiality will be maintained
- instructions on how to complete the questionnaire, including information on rating scales and "no answer" situations
- typical rating errors.

We've all heard promises about

the era of the paperless office, but that time still seems distant to many of us. The paper-and-pencil solution (often machine readable) is by far the most popular means of survey administration, with every indication that its popularity is stable, if not increasing. This format is easily accessible and familiar to almost everyone.

The greatest liability of the paper-and-pencil method, when compared to alternative technologies, is the turnaround time it requires for mailing and processing. Most multirater processes can accommodate these time requirements. But organizations that need to see survey results right away will find paper surveys to be too slow, unless they process the information on site.

Electronic data-collection methods certainly ease data processing. But having a paper version of the questionnaire as backup can be invaluable. In many instances, a paper version has allowed a processor to recover data that was lost due to miscoding during administration of the survey.

Some organizations have turned to on-line, telephone, or fax methods of administration for multirater instruments. Each method has its own advantages and disadvantages. Whichever technology you are considering, ask yourself the following questions:

- Do all potential raters have equal access for providing input?
- Are raters comfortable with the technology?
- Can you guarantee confidentiality well enough to satisfy the raters?
- Is the length and content of the questionnaire compatible with the proposed technology?
- If you are changing administration methods or using more than one method, could you be creating response biases? Are results really comparable across technologies?

The effect of method on the responses is unclear. Until that ambiguity is resolved by research, it's probably best to pick one administration technology and stick with it.

One of the most critical steps in the total multirater process is to ensure that the identity of the person being rated is accurately coded and captured on the form—if for no other reason than to let the rater know whom to rate! This, of course, becomes doubly critical during processing, when administrators must be certain that the right raters are included in every ratee's report.

The least preferred method for coding a questionnaire is to let the respondents do it themselves. Experience shows that, regardless of instructions given, a lot of people are either unable or unwilling to record accurately the name, and usually a code number, of the person they are rating.

(Note that the name alone is generally not enough. That's because of duplicate names within an organization and because of the inevitable preponderance of nicknames, initials, recent name changes, and so on, which can result in one person being identified by several different names.)

The next most preferred method is to require the manager to code the questionnaires with the required information, such as his or her name and employee number, before distributing the forms to the raters. This is a relatively cost-effective method of coding questionnaires, which places responsibility for accuracy in the hands of the ultimate consumer of the information—the manager.

But human beings are fallible; we make mistakes that can result in inaccurate coding. And we procrastinate, leading to delays in the distribution of questionnaires. That leaves the raters with inadequate time to complete the form, which makes them more likely to miss the deadline for returning it.

The most accurate way to code forms is to have them precoded—during printing—with key information on the manager being rated. This won't work unless the data used to generate the codes are accurate and complete, a requirement that, unfortunately, taxes some current human resource information-systems groups. If designers create the form to be optically scanned, they can have the manager's code printed on the form for capture during the scanning process, further ensuring accuracy.

Be sure to let everyone know that the preprinted codes indicate the

Why More Companies Are "Asking Around" for Feedback

The use of "full-circle" feedback and other kinds of multirater systems is surging; more and more companies are "asking around" for input on employees' performance—input from subordinates, peers, supervisors, managers, and even customers. Several factors contribute to the trend's increasing popularity in organizations:

- Multirater systems complement other popular organizational initiatives, such as empowerment, employee participation, organizational flattening, and teamwork.
- Most traditional performance-appraisal systems emphasize the role of an employee's manager in performance management. Such systems are increasingly insufficient as feedback and development tools, because downsizing has increased managerial spans of control. In other words, managers who supervise 15 employees or more are unlikely to have the time and knowledge to appraise and develop all of them effectively.
- Other traditional means for creating organizational change through employee input—for example, through the use of employee-attitude surveys—continue to fall short of expectations. Frequently, this is due to inadequate follow-through and a lack of accountability for subsequent action.
- The creation of a new multirater instrument gives management the opportunity to "operationalize" a vision for success, through descriptions of behaviors that support organizational values.
- Multirater systems "feel" more reliable than single-rater feedback systems. Multiple raters can provide a variety of perspectives; people generally assume that those viewpoints will add up to an accurate assessment of an employee's performance.

ratee only—not the rater! You can even reassure skeptics by telling them to feel free to trade blank questionnaires within their work group—but not with employees who are rating other managers.

Processing and reporting feedback

Most multirater processes are characterized by a large number of raters providing information on a large number of ratees who expect timely, accurate reports on the results.

The need for rapid and totally accurate reporting in large volumes provides a significant challenge for the data processor. Most companies find the best way to meet that challenge is to hire an external supplier with the experience and resources to do the job. The use of an outside supplier also helps assure raters of anonymity. Of course, many internal processing functions can and do maintain confidentiality for data that are just as sensitive. But a lot of employees feel safer dealing with an outside group.

During processing and reporting, make every effort to capture accurately the responses of the raters and to link them with the correct ratee. One solution we have used successfully involves a combination of bar coding and optical-scanning technology.

The planners of any multirater system will at some point have to decide on a minimum number of raters who must turn in responses before a report will be generated on a ratee. It's important to set a minimum number, in order to maintain raters' anonymity in case only one or two raters turn in responses about a particular ratee.

Probably the most commonly used minimum group size for employee surveys is five respondents, meaning that no data is reported for any item that doesn't have responses from at least five raters. Five has a certain appeal, but the reality is that many managers have only a handful of direct reports. Requiring five responses would prevent a lot of managers from receiving valuable feedback. Even with five or more direct reports in a work group, a very high response rate might still be needed before a report could be generated for a manager.

Many companies have settled on a minimum group size of three respondents to create a report. A handful of multirater processes allow data to be reported on only two respondents—or even on one respondent. Of course, it's impossible to maintain a rater's anonymity if he or she is the only respondent, and it's almost as difficult to maintain anonymity with two respondents. So if anonymity is any concern at all, it's best to require at least three responses in order to generate a report.

Because some managers have only a few people reporting to them, it is crucial that every completed form is received and processed in time to be included in the report. System planners should have procedures in place to encourage quick responses and to ensure timely processing, whether the responses come through the mail, by phone, by fax, or through on-line systems.

Resources on Performance Appraisal

Need more information on performance feedback? Try the following articles that have appeared recently in *Training & Development*.

- "The Power of Peer Review," by Martin L. Ramsay and Howard Lehto. July 1994.
- "Give It to Me Straight," by Jerry Baumgartner. June 1994.
- "360-Degree Feedback: The Whole Story," by Kenneth M. Nowack. January 1993.
- "How To Do Peer Review," by Gloria E. Bader and Audrey E. Bloom. June 1992.
- "Upside-Down Performance Appraisals" (in "Four by Four,") by Catherine Petrini. July 1991.

To purchase reprints of these articles, please contact ASTD Customer Service at 703/683-8100. Use priority code BVM. Single-article photocopies are $6 each; the package is $15. Phone for prices on bulk orders (50 or more) of custom reprints. Call to order by credit card, or send your order to ASTD Customer Service, Box 1443, Alexandria, VA 22313-2043.

Feedback and action planning

One of the appeals of multirater processes is the accountability that they establish for the person being rated. A well-constructed rating instrument will include items that meet the following requirements:

- They relate specifically to the ratee, with no confusion as to who is being evaluated.
- They cover topics that the ratee can control.
- They relate to behaviors and processes that the ratee can act on and that are easily translated into a development plan.

Once the data are in, the ratee must be able to translate the results into a coherent action plan. To do this, the ratee will need ready access to the information, in enough detail so that he or she can put it to productive use. Ratees must also have the skills and support needed to analyze and interpret the results. And the organizational environment should not only require action to occur, but should also provide the necessary resources for making it happen.

Approaches to achieving those objectives will vary, depending on the size of the population requiring support, the amount of experience the rated managers have with using multirater information, and the resources available to the organization. Three possible solutions—listed in order from the most "high touch" to the least personal—are facilitators, workshops, and workbooks.

Depending on the number of managers who need support and the amount of resources devoted to the process, trained facilitators can be invaluable. This is especially true when an organization first begins using a multirater system. At such times, anxiety tends to be high, and managers' skills in productively using the collected information may be low.

Facilitators can handle the following tasks:

- delivering reports to the targeted managers in personal, one-on-one meetings, in which the facilitator shows the manager how to read and interpret the report
- helping managers develop and write action plans

- facilitating the presentation and discussion of the results with raters
- establishing a method for ongoing review of ratees' progress against their action plans, particularly in multirater systems in which the use of the feedback is for purely developmental purposes.

Workshops provide a group setting for personal interaction and mutual support of manager/ratees. A trained workshop leader can assist the managers in the proper and full use of the feedback. She or he can also communicate information regarding the company's position on the proper use of the information and the commitment to the process. In addition, such a workshop—typically a half-day long—can provide the managers with valuable time for quiet, uninterrupted analysis of their reports, away from the competing priorities of their jobs. This benefit should not be discounted or underestimated.

Practicality is a common argument in favor of workbooks as the chief means of supporting a multirater system. But, as Ayn Rand says in *Atlas Shrugged*, "The evaluation of an action as 'practical' depends on what it is one wishes to practice."

On the other hand, the reality is that many multirater systems are large-volume processes that are conducted in multiple locations. Their scope simply does not permit the full-scale implementation of workshops or facilitator networks.

In such situations, a workbook is a minimum requirement. Typically distributed along with the data reports, the workbook assists a manager in analyzing the report and creating a way to distill the key strengths and opportunities into a development plan. Workbooks also serve to standardize the "output" of the process (the action plans) across managers.

Workbooks could also include the following useful items:

- formats and overheads for conducting feedback meetings
- suggestions on how to conduct feedback meetings (dos and don'ts)
- suggested time lines for events
- listings of internal and external training resources tied to targeted skill areas.

Use of multirater feedback

If multirater feedback is to be incorporated into personnel systems for use in decision making, then consistent and accurate use of the data is crucial.

Most multirater systems use a category structure for reporting feedback, with each category encompassing a set of survey questions. Category scores are easier to use than individual items; in fact, some consultants recommend reporting only category scores. Also, category scores should be more reliable than scores on individual items.

But "armchair" categories derived from the arbitrary clustering of items don't always meet the requirements of good measurement. Base your category structure on a statistical (or factor) analysis of the item relationships. For the analysis, use data from a pilot survey administration—or, at least, data that have been confirmed after the first round of data collection. Note that a factor analysis can have the added benefit of identifying questions that are unclear or redundant. Eliminating redundant questions can shorten the questionnaire, possibly raising response rates.

> "THE EVALUATION OF AN ACTION AS 'PRACTICAL' DEPENDS ON WHAT IT IS ONE WISHES TO PRACTICE"

Once data processors have calculated and reported category scores, they then must integrate the information into the personnel system. The most typical question is how to use the information in a performance appraisal in a way that is valid, consistent, and fair. Appraisers' jobs will be easier if the multirater feedback categories directly correspond to the performance-appraisal categories, or rating dimensions, that the organization uses. Ideally, those rating dimensions will be listed on the performance-appraisal instrument itself.

Provide appraisers with explicit instructions on how to use the multirater feedback in the appraisal, including how to factor it in when determining performance-appraisal ratings. Wide disparities in how different appraisers weight the multirater data can create confusion, dissatisfaction, and legal problems.

One solution to this problem is to use the multirater information to create a development action plan for each manager, which is recorded in the appraisal process. In subsequent appraisals, the manager is evaluated on his or her performance in achieving the action-plan objectives.

This process further promotes the joint responsibility of the targeted manager and her or his boss in creating developmental opportunities.

Embarking on the multirater journey

Some organizations make the move to multirater feedback systems because the concept "feels right." But don't lead your organization into blindly implementing such a system without considering all the relevant issues and questions. We know far too little about the characteristics of these systems. And our lack of knowledge is compounded by the widely varying methods used by different organizations.

We can draw from a wealth of experience gained from other assessment processes, but our understanding of multirater systems is just beginning. So keep your eyes open, and get expert help from internal and external sources as you embark on your multirater journey. ■

David Bracken *is the director of organizational assessments and research services for Personnel Decisions, 2000 Plaza VII Tower, 45 South Seventh Street, Minneapolis, MN 55402-1608; 612/337-8289.*