

The 360-degree Dilemma

Overcoming the challenges of using multi-rater assessment for performance appraisal.

By Tracy M. Maylett and Juan M. Riboldi

The use of multi-rater feedback, also known as 360-degree feedback, has increased dramatically over the past two decades. Some estimates suggest that as many as 90% of all Fortune 500 firms use some type of multi-rater feedback with their managers. However, the purpose behind these assessments varies greatly.

Originally used almost exclusively for developmental purposes, 360-degree feedback is seeing increased use for purposes other than employee development. One application that is gaining increasing popularity is the use of multi-rater feedback for performance appraisal. Opinions, however, are mixed as to whether multi-rater feedback should also be used for appraisal purposes, or used only for development. Where two very opinionated camps

once existed on this issue, the lines are beginning to blur.

We often see multi-rater feedback used for both development and appraisal purposes. However, there are important differences between the two. Understanding and designing

the feedback process to account for these differences could determine whether your formal evaluation processes succeed or fail.

Why Feedback?

Before determining whether to use multi-rater feedback for development or for appraisal, it's important to understand the role of feedback within an organization. Feedback is a vital part of performance, growth, and development. Understanding ourselves, and how we interact with others, helps us to understand what impact we have on those around us. The perceptions of others within our circle of influence—whether those perceptions are accurate or inaccurate—determine, to a large degree, our level of success. Regardless of the accuracy of these perceptions, our interaction with others both influences and is influenced by the perceptions of others. This is where 360-degree feedback comes into the picture.

Based on the philosophy that individuals should receive a full 360-degree picture of performance by gaining multiple perspectives, multi-rater feedback gathers input about an individual's performance by soliciting feedback from those stakeholders impacted by that individual. Similar to the 360 degrees of a circle, an individual is



figuratively at the center of that circle, and feedback is gathered by way of a survey from those in positions to observe the person's performance: supervisors, direct reports, peers, customers, etc.

Multi-rater Appraisal

As many companies saw the advantages that multi-rater feedback carried within the employee development

programs, many quickly perceived 360-degree feedback as the panacea for their appraisal woes. There appears to be the added pressure of wanting to "get the biggest bang for the buck." Using 360-degree feed-

back for performance appraisal, as well as development, appeared to be a cost effective solution. The question from many executives soon became, "Why not just combine the two?" However, most managers failed to realize that the purposes behind these two forms of feedback—although similar in design—produce significantly different results.

First of all, it is important to note that we typically recommend that an organization use, or at least start off by using, multi-rater assessments for development purposes only. Because many organizations (and managers) operate in feedback-poor environments, first exposure to a multi-rater assessment may be accompanied by some degree of angst on the part of both the organization and the employee.

When employees understand that the feedback they receive will be used purely for their own developmental benefit, they tend to be more receptive to the feedback provided. Rather than receiving the feedback from a defensive posture, they are more apt to accept the feedback as a "gift" from those they impact.

When first exposure to multi-rater feedback has administrative consequences (i.e., raises, promotions, bonuses, possible layoffs, etc.), however, employees may be more likely to attack the feedback, rather than accept it. After all, they may have much more to lose. Similarly, raters are less likely to provide honest, candid feedback when they know that it may impact someone's pay or position.

While we encourage the use of multi-rater feedback for development, we also realize that reality says that often the decision to use 360s for appraisal has already been made. We have seen a number of organizations successfully use 360-degree feedback for development, then begin using it for appraisal purposes at a later date, once the organization becomes more adept at giving and receiving feedback. These companies often find that gathering multiple perspectives for appraisal provides very rich data.

We have also seen other organizations successfully use multi-rater appraisal from the very beginning. The organizations that succeed are those that understand the difference ahead of time, and build these into their processes.

Appraisal Woes

A significant percentage of companies report that their traditional performance appraisal process—a top-down, supervisor-to-subordinate process—is inadequate and fails to provide an extensive view of performance. Other organizations find that their performance appraisal metrics generally do not take into account the manner in which employees go about reaching targets.

We recently worked with a group of senior leaders that were regularly achieving revenue targets. Because of this, these managers consistently received exceptional performance appraisal ratings. However, they also experienced high levels of employee attrition. They had difficulty attracting and retaining talent. It soon became clear that these managers were meeting short-term targets at the expense of long-term profitability. They were also leaving a "wake of dead in their path," as the company described it, in achieving these short-term revenue targets. Although their performance appraisals rated them as excellent performers, these managers and their departments were often found to be the origination points of many of the employee issues.

It was quickly realized that the top-down appraisal being used by this organization only took into account the ratings of the supervisor. These ratings were based almost entirely on whether or not the manager hit monthly revenue targets. However, the managers' true overall performance was largely unmeasured. The company made the decision to use multi-rater feedback to look at overall performance and leadership behavior. Through this process, they found that many of these so-called "stellar performers," were actually not so stellar after all.

Little, if any, correlation exists between a person's traditional (top-down) performance appraisal and the same person's developmental multi-rater feedback.

In fact, it became apparent that their traditional performance appraisal was ineffective at determining true levels of performance—particularly in terms of leadership behaviors.

Multi-rater Advantages

Multi-rater feedback has several significant advantages over single-rater assessment. Rather than relying on the perceptions of one individual, multi-rater feedback takes into account multiple perspectives. This is especially critical when one person (i.e., the supervisor) does not have the opportunity to observe all areas of an employee's performance. Those working with the employee, along with the supervisor, are generally able to provide a more comprehensive picture of an employee's behavior or performance.

The question is also often raised as to whether managers are skilled (and impartial) at providing feedback through performance appraisal. While some organizations provide appropriate training to supervisors in how to evaluate employees, most do not. Not only does this raise questions as to accuracy of performance appraisals (which is a constant source of employee complaints), it can also surface legal concerns.

Appraisal vs. Development: A Big Difference!

It is critical to be clear about the purpose behind an assessment. Is the purpose of the process administrative/appraisal focused, or is it intended purely for development? Failure to clearly answer this question up front could spell disaster.

We recently completed a comprehensive series of studies that show that little, if any correlation exists between a person's

traditional (single-rater, top-down) performance appraisal and the same person's developmental multi-rater assessment. Additional studies show that when an employee's developmental (positioned as solely for development purposes) 360-degree feedback score was compared to his or her performance appraisal score, there was little correlation. We might expect that those receiving very high 360 scores would also receive the highest performance appraisal scores. Those in the middle ranges of one would naturally be in the middle ranges of the other. These studies found that this was not necessarily the case. However, those who were the highest performers (top 10%) on a developmental multi-rater assessment typically scored in the top half of performance appraisal scores when compared to others. Similarly, those who ranked in the bottom 10% of multi-rater scores were generally ranked in the bottom half of performance appraisal scores.

The 'What' vs. the 'How'

A first glance at these studies bring into question the validity and reliability of 360s to begin with. While it is true that validity and reliability of the instruments are important, the primary issue here is not one of statistics. It lies in purpose of the instrument. The primary purpose of a performance appraisal is to rate an employee's performance based on *what* he or she has accomplished. It measures skills, performance, and accomplishment according to established

metrics, goals, and company or unit objectives. The purpose of a developmental assessment is to measure *how* it was accomplished, and what behaviors brought this about. In other words, theoretically, a manager could accomplish all of his or her stated objectives: something that is measured in a performance appraisal. This is the *what*. However, the manner in which he or she accomplished it is measured in a developmental assessment. This is the *how*. He or she may have hit all the key performance goals, yet destroyed important relationships in his or her path (haven't we all worked for this manager at some point in our careers?). Because of this, it is possible to receive high scores on one form of assessment and low scores on the other.

Absolute vs. Relative

One of the concerns typically expressed about multi-rater feedback is that the "scores are too high." On a 5-point scale, this may mean that the scores hover around the 4.0-4.2 range. When you stop to examine this, it is no surprise that this is the case. The vast majority of employees should be performing at this level. If not, the company has a separate performance issue. However, it is important to recognize that, by nature, multi-rater feedback used for development will typically result in higher (by about 20%) mean, median, and mode scores than do performance appraisals.

Most managers fail to realize that the purposes behind these two forms of feedback—although similar in design—produce significantly different results.

In fact, over 90% of employees will receive higher overall developmental feedback scores than they do on their overall performance appraisal scores when rating scales are similar.

Part of the reason for this is that those providing feedback for development tend to look at *absolute* performance, rather than *relative* performance. In other words, on a scale from *A* to *Z* in terms of how the employee *actually performed*, where did the employee fall on an *absolute* scale? On the other hand, when used for appraisal purposes, raters tend to look at the *relative* performance—how well did they perform relative to others? It's a bit like the "grading on a curve" that we all came to know and love in school.

This also means that a greater range in scores will likely be found in feedback used for appraisal versus feedback used for development. This is a fact that tends to encourage proponents of 360s for appraisal; there is a greater spread in scores between high scorers and low scorers. However, it is also important to note that much of this variability is the result of a wider range in scores provided by supervisors, and not necessarily the other raters. Much of this comes from pressure on the supervisor—real or perceived—to make a visible distinction between high and low performers.

This pressure is often a carry-over from the need to provide clear performance ratings in order to divide up bonus or compensation pools. Unfortunately, this forced-ranking system may cause supervisors to spread scoring ranges out to the point that they are no longer indicative of true performance levels, and therefore are of questionable use to the person receiving the feedback report.

Who Owns the Data?

This also brings up another key difference—who owns the data? Typically, when used for development the person being rated "owns" the data. In other words, the data is generally presented to the employee first, often with great concern for complete confidentiality. The employee is often the only person to see the data, unless data is (willingly) shared with a supervisor. Occasionally, Human Resources has access to the data as well, but not always. The upside of this is that there is a perceived safety net. The employee knows that the data is purely developmental. The downside, however, is that development is left completely up to the employee—which may or may not lead to change.

When used for appraisal, on the other hand, the employee's supervisor, as well as Human Resources, typically have access to the data either before or at the same time as the employee. Quite often this leads to sources other than the employee "owning" the data. This often leads to greater accountability.

Perception Gaps

Another important difference in using multi-rater assessment for appraisal versus development comes to light when examining perception gaps. When used for development, identifying gaps between raters' perceptions is one of the most important elements of the feedback. If I see myself differently than others see me, this gap in perception is a valuable source of information. Similarly if there are gaps in the way my supervisor rates me versus the way my peers rate me, this is critical information.



Generally with development, employees are encouraged to explore these gaps and ask questions of clarification from their raters in order to understand the reasons behind the ratings. When used for development, these conversations are generally very productive and often fairly comfortable. However, these gaps in perception, particularly in supervisor ratings versus others' ratings, are somewhat distorted when used for appraisal (as discussed previously).

Gaps between raters' self-ratings and supervisor ratings are often significant—particularly in appraisal. If this is difficult to believe, you may want to try a little experiment. Ask a room full of people on the same work team to close their eyes and raise their hands if they consider themselves to be below-average performers. Chances are very few hands will go up—an interesting observation, as statistically half of the room is at or below average. Now, ask the supervisor about the same group of individuals, and he or she will likely be able to tell you which half of the group is below average. This difference is exaggerated through performance ratings.

Because of factors such as these, gaps in scores may not provide the valuable information they do when used for development. Additionally, asking for clarification of ratings from ratees may not be as appropriate in appraisal feedback, and may often be perceived as defensive or even hostile.

Making it work

A number of organizations have successfully used 360-degree feedback for performance appraisal. These organizations appear to share several common characteristics that help them to succeed where others fail:

- ***Understand the differences in use and purpose.*** These two types of assessment are not interchangeable. Understanding that scores will differ depending on the purpose, as discussed previously, will help in determining how best to use and interpret the scores. As outcomes will likely be different depending on purpose, these differences should be taken into account when determining how best to use and interpret the results.
- ***Communicate the purpose and process.*** Let employees know the intended purpose before administering the assessments, as well as how the results will be used. Communicate the process and hold to it. Train the organization on how to provide accurate ratings.
- ***Use a pilot group.*** Using a pilot group (or groups) of 35-50 people prior to organization-wide rollout has several advantages. First, it

allows for refinement of the process and of the instrument itself. Many potential problems are quickly identified through this process that would have been widespread otherwise. Second, the pilot group can act as champions throughout the rest of the organization. This pilot study may also give you a small taste of what to expect when rolled out to the rest of the organization.

- ***Wait before taking administrative action.*** Although these successful organizations use 360s for appraisal, most have waited 12-18 months before beginning to tie raises, promotions, etc. to the appraisal results. This allows people to become familiar with the process and comfortable with providing feedback.
- ***Select appropriate raters.*** It is often more appropriate for employees to select their own raters with developmental feedback than with appraisal feedback. With appraisal, there may be the temptation to “stack the deck” in terms of who is selected to provide feedback. This can be solved by selecting raters on behalf of the employee, rather than requiring the employee to select his or her own raters. It is also critical to ensure that selected raters have regular interaction with the employee being rated, and can provide accurate feedback as to performance. It is also important to limit the number of raters.
- ***Design the appropriate survey.*** Multi-rater assessment for development should include questions

geared at behavior (the *how*), while appraisal assessments can focus more on the performance (the *what*). Survey questions should reflect these differences in design. We also recommend that a 7-point Likert scale be used, rather than a 5-point scale. This allows for greater differentiation in scores. It is also important to design a survey that is short enough that it can be completed in 15 minutes. We have found this to be approximately 45-55 questions. Any longer and raters tend to experience rater fatigue, often resulting in *all* scores falling in the “good” range.



When using these assessments for appraisal purposes, it is also important to consider the number of people that will be involved in providing feedback. Multi-rater appraisals involve more of the organization in terms of providing feedback. Each person (especially managers), may be required to complete multiple evaluations. The shorter the survey, the more apt they may be to provide the time needed to give accurate information. It therefore becomes especially important that the survey be even more concise, as completing multiple, lengthy assessments may become time-

consuming and ineffective. Keep in mind, however, that this brevity should never be at the expense of losing the value of comprehensive feedback.

Be aware that most off-the-shelf surveys are designed for developmental use, not for appraisal. We typically find that off-the-shelf surveys are inappropriate for appraisal purposes, and that organization should consider a survey customized to their specific purposes.

- **Don't group questions into single category scores.** Many performance appraisals group a series of questions into one question. In other words, only one score is given to an overall category. An example of this would be the category of "Communications." A "Communications" competency is comprised of many elements: oral communications, written communications, listening, etc. Rather than providing scores for each question, many appraisals will give one overall score for a category (such as "Communications"). When this is the case, it is often difficult for an employee to know which area of "Communications" is being addressed. Providing a score for each individual question provides more useful data, and is easier for the rater to evaluate more accurately.

Answering the Question

So, the question still stands as to whether or not 360-degree feedback should or should not be used for appraisal purposes. Ultimately, it depends on the needs of the organization. Using multi-rater assessment for appraisal

has some significant advantages, as well as disadvantages. It is important to note, however, that 360s for performance and 360s for development are likely to produce different outcomes. Also, multi-rater feedback used for development shows little correlation to traditional top-down appraisals. They are not interchangeable, and should not be treated as such.

Ideally, we recommend that organizations use *both* a multi-rater assessment for development *and* a separate performance evaluation, and that they be clearly separated.

When both instruments are used together, the developmental piece can focus specifically on the *how*, identifying specific leadership behaviors, rather than goal accomplishment. The evaluative piece then focuses more on the *what*, including such traditional elements as achievement of performance targets, MBOs (management by objectives), and organizational metrics.

Using both instruments will provide a more complete picture of overall performance. Both make up important elements of an employee's overall development plan. Whether or not the appraisal piece comes in the form of multi-rater feedback is something that requires careful consideration and design, as it does pose some unique challenges. Used effectively, multi-rater feedback has the potential

to provide valuable insight for the organization and the individual.

Organizations that are successful at using multi-rater feedback for performance share a few common success factors. These organizations:

- Understand the differences in use and purpose.
- Communicate the purpose
- Use a pilot group.
- Wait before taking administrative action.
- Select the appropriate raters.
- Design the appropriate survey.
- Don't group questions into single category scores.

Tracy Maylett and Juan Riboldi are Sr. Partners at DecisionWise, Inc., a talent and organization development firm specializing in turning feedback into results. Tracy holds a doctorate in Organization Change, and Juan a Masters Degree in Organization Behavior. Each has worked with numerous organizations across the globe over the past 15 years. Tracy and Juan have taught on the Faculty of several Universities, including Brigham Young University.



DecisionWise, Inc.
Tel: 1-800-830-8086
www.decwise.com