Should We Bother With Performance Appraisals?

Depends on whom you ask. We put the question to two experts. Their subsequent e-mail dialogue scrutinizes the merits (and demerits) of this much-beleaguered corporate practice.

By Dick Grote and Steve Scullen

INTRODUCTION

Nobody likes performance appraisals—neither the manager doing the evaluating, nor the employee being evaluated. Clearly, evaluations are front-runners for being the most maligned of corporate processes. And yet they should be a good thing. After all, doesn’t everyone need—and deserve—to know where they stand? In theory, yes, but bring theory to ground and it’s quite a different story, often one with a sour aftertaste if not a bad ending.

We know there is something wrong with the performance evaluation-process, which is why we’re always trying to fix it—360-degree feedback and forced ranking are two of the more prominent recent fixes. But none of the fixes seems to satisfy everybody.

Are we expecting too much—or the wrong thing—from performance evaluations? Can they be made fairer, less intimidating, a better experience for all concerned, with better results? This was among the questions in an e-mail dialogue between two men who, while not on opposite sides of the fence, have different takes on the value of many aspects of performance reviews:


• Steve Scullen is associate professor of management at Drake University and author or co-author of several journal articles on performance appraisals and forced ranking.
Greetings, Steve!

To kick off our dialogue on performance appraisals, I’ll fess up to an obvious fact: Most managers hate conducting performance reviews. If they thought they could get away with it, they’d probably skip the whole annoyance completely. In too many places, performance evaluations are sloppily done and not taken very seriously. A lot of supervisors would rather endure a colonoscopy than write and deliver a performance review, particularly if there are some hard, cold truths that they can’t avoid discussing.

But in spite of all the problems and resistance, I’m a solid believer in performance appraisal. I think performance appraisal is critically important for any organization that’s more sophisticated than a mom-and-pop store—or that wants to be.

As hard as performance appraisal may be—and done right, it is hard—I’m convinced that we do it because it’s an ethical obligation of leadership. Every person on the team wants the answers to two questions. First: What do you expect of me? Second: How am I doing at meeting your expectations? The performance-evaluation process answers those two questions. So why do we do performance appraisal? Because as leaders we have a moral obligation to do so.

And we also have an obligation to put the time into performance appraisal that it deserves. Goodness gracious, Steve! Managers don’t spend a tenth as many hours assessing and developing and appraising people as they spend in the restroom. But they’ve got the gall to whine about appraisal taking too much time. That’s nonsense.

Best regards!
Dick

From: Steve Scullen
To: Dick Grote

Hi Dick—

Yes, most managers do hate doing performance reviews, and most of them spend too little time working on them. I also agree that most people want to know what is expected of them and how well they are meeting those expectations.

But I have concerns about using formal performance-appraisal systems as the means for delivering that kind of information. I say that for a couple of reasons. First, I think that the discussion of expectations and the associated dialogue, coaching, mentoring, etc., should be an ongoing process that begins long before the performance review itself. Ideally, the formal performance appraisal would be largely a matter of formalizing all that has gone into that process. The communication of expectations and progress toward meeting those expectations could be accomplished without a formal performance review.

Second, a lot of research (some of it mine) has shown that performance appraisals have disappointing levels of reliability and validity. Without getting into technical details, I can say that it calls into question the accuracy of appraisals in general and therefore the soundness of any decisions that might be based on them.

You indicated that if managers do performance appraisals correctly, the payoffs will come. What are those payoffs? And do we really need appraisals to realize those payoffs?

Best,
Steve

From: Dick Grote
To: Steve Scullen

Greetings!

There’s no question, Steve, that the discussion of expectations and coaching and feedback should be a routine and ongoing process. And some managers—the great ones, the naturals—do these things routinely. But managers such as this are few and atypical. Left to their own devices, too many people with manager titles will give the goal-setting and feedback requirements short shrift. They’ll stick it on the “when I get around to it” pile.

That’s why we need a formal process. Sure, the ideal is to have an environment of constant feedback and communication. But an environment like that is rare. That’s why it’s important to require managers
to, at a minimum, discuss goals and expectations
with each staffer at the start of the year, and then, at
year’s end, evaluate how well that person has done.

Reliability and validity are also worthwhile con-
cerns, but you academics tend to make a much big-
er deal of this issue than we practitioners do. No
argument—reliability and validity are important. But
companies aren’t psychology research labs. The
question to which every employee wants the answer
is, “Boss, how am I doing?” While on an organization-
wide basis the overall reliability and validity may not
meet academic standards, on an individual basis, if
people are getting the straight scoop from their boss
on their performance, that’s OK.

But there is some good news here. Many compa-
nies actually are concerned with increasing reliability
and validity and are setting up calibration meetings
or “rater reliability” sessions. In these meetings, su-
pervisors review with each other the performance-
appraisal ratings they’re planning to give their subor-
dinates before they actually hand them out. They
check each other’s intended ratings before they’re
made final in order to assure that people who work
for different supervisors are held to similar standards
and that there’s a level playing field for all.

Best regards!
Dick

---

From: Steve Scullen
To: Dick Grote

Hi Dick—

Sure, employees want to know how the boss
thinks they’re doing. And if the boss gives them the
straight scoop, through whatever channel, then they
will know that. The employee is well served by hav-
ing that information, in that it allows her the oppor-
tunity to bring performance into better alignment
with what the boss expects.

But if I were boss, I would want more than that
for my organization. Truly valid appraisal information
promotes good decision-making. Employees can use
it to make appropriate adjustments in their job per-
formance, and management’s ability to make optimal
personnel decisions should be enhanced. Unfortu-
nately, however, both benefits are reduced by any
invalidity that creeps in (and I find the validity evi-
dence disappointing). What if employees make the
performance changes their bosses were looking for,
but those changes don’t truly serve the organization?
Or what if faulty appraisal data leads management to
make ineffective personnel decisions? It almost cer-
tainly follows that the organization won’t perform as
well as it could.

The rater-reliability sessions you mention are a
good step toward greater validity. This is important
not for academic or theoretical reasons but, rather,
because organizations and their employees need
valid information to perform up to their capabilities.

Still, I have doubts about whether the formal
performance-appraisal systems we see in practice
today hold great promise for improving our organiza-
tions.

Best wishes,

Steve

---

From: Dick Grote
To: Steve Scullen

Greetings, Steve!

So far we appear to concur on a couple of key
issues. First, we agree that getting accurate perform-
ance information to people is an important ethical
responsibility of leadership, even though a lot of
managers don’t like the process their company asks
them to use. Second, we both realize that there are
some bumps with the way the system works. But my
response to attacks on performance appraisal’s
shortcomings is the same one Bill Clinton used when
defending affirmative action: “Mend it, don’t end it.”

Here are some ways we can mend it. First, com-
panies must raise their expectations about the quality
of appraisal execution, setting their minimum stan-
dard for getting appraisals done at 100 percent un-
complaining compliance. If a manager’s sluggishness
in completing an appraisal means that a subordi-
nate’s merit increase is delayed, HR shouldn’t let her
off the hook by automatically making the increase
retroactive. She must take the heat of explaining to
Sam why her slothfulness in getting his appraisal
written caused him to miss out on a salary increase
for a pay period or two.

Second, top executives must threaten significant
consequences if managers don’t take the process
seriously. Hunt Oil’s performance-appraisal process, for example, requires every manager to discuss thirteen open-ended performance-related questions with each subordinate every March. The only writing the system requires is a memo each manager has to send to CEO Ray Hunt by March 31, saying either that she has conducted all her discussions, or that she hadn’t done so along with an explanation of the reason why. And that explanation had better be good, since on April 1 Hunt picks up the phone and starts calling. As Hunt’s VP of HR explained, “You don’t want to get that call from Ray Hunt.”

Another great way to build accountability is to insist that senior managers review and sign off on all performance appraisals before the supervisor who wrote them can sit down and discuss them with her subordinates. This review requirement not only helps assure timely completions—it boosts inter-rater reliability by making sure that people who perform at the same level of quality will get the same appraisal rating, whether the appraisal is written by manager A, manager B, or manager C.

Most important, the reviewer—the senior manager—can guarantee that tough-minded, demanding performance standards are set. Some managers (like some college professors) are more lenient than others. The reviewer can make sure that the standards of the department’s toughest appraiser set the bar for everybody else. That’s the best way to counteract the Lake Wobegon effect, in which everyone is rated as Exceeds Expectations.

Best regards!
Dick

From: Steve Scullen
To: Dick Grote

Hi, Dick—

I agree that companies need to raise their expectations about the quality of appraisal execution, and that executives should demand that all managers—including themselves—take the process seriously. Managers would put a good deal more effort into that part of the job if it were clear to them that executives value high-quality appraisal and are willing to reward managers who develop and deliver them. Unfortunately, survey research I’ve seen indicates that the majority of managers say they are not rewarded for that kind of thing. It’s little wonder this is a low priority for many managers.

The preceding assumes, of course, that managers could deliver quality performance appraisals if they put enough effort into it. I am convinced that most could do better than they do now, but the question of just how well they might do is unsettled. Evidence from 360-degree feedback systems (which I’ll admit is not exactly the same thing as performance appraisal) consistently shows that the differences in ratings across ratees are smaller than the differences across raters—that is, managers are using completely different criteria to judge people. Ideally, all raters would give exactly the same rating to a particular ratee. But ratings tend to be more a reflection of the one who gave the rating than of the one whose performance was rated. That might be OK if it were happening because each rater sees—and accurately rates—somewhat different pieces of a ratee’s total performance. I’m sure this happens to some extent. But the ratings variance attributable to raters is considerably greater than the variance attributable to ratee job performance. The dominant factor in ratings should be ratee performance, but it is not!

Maybe it doesn’t have to be that way. It could be that with the proper interventions, raters could give more consistent and valid ratings. Using multiple raters, assuming they are knowledgeable about a particular ratee’s performance, holds promise. Much of the idiosyncratic variance in individual raters’ points of view is eliminated if ratings are averaged across raters. Maybe increasing the importance that executives attach to appraisals would do the same. I would hope so, but I haven’t seen empirical evidence to that effect.

I’ll wind this up by stating the obvious: Regardless of whether formal appraisal systems are used, managers will form opinions about employees’ job performance, and personnel decisions will be made in accordance with those opinions. Eliminating appraisals wouldn’t change that. So I am not averse to the idea of performance appraisal per se.

Best wishes,
Steve
Greetings, Steve!

I’m delighted that you’re not averse to performance appraisals; no one seriously could be. And of course we should always try to improve our organizational processes. Here’s how to improve performance appraisal.

First, we need to forget about 360-degree feedback, at least for performance-appraisal purposes. Certainly, 360-degree feedback may have a place—a minor place—in helping people get a better understanding of their development needs. But it has no place in conventional performance appraisal. To allow anonymous employee assessments into part of the formal evaluation tool does more than just encourage biased and self-serving responses—it poisons the entire well in terms of the original objective. It is particularly inappropriate to tie pay, promotions, development opportunities, and terminations—the things that a strong appraisal system controls—to anonymously provided assessments. The issue is not whether underlings and co-workers can provide relevant information. They can. The issue is whether they should be allowed to do so in a context where they cannot be held accountable.

What we need to do, Steve, is add forced ranking to our performance-management processes. Forced ranking requires senior managers to look over the organizational talent pool and, based on their performance and potential, identify the organization’s top talent (the A players), the solid-performing middle (the B players), and those bringing up the rear (the C players). Forced ranking can drive the truth into performance management, since not only does it force managers to identify the organization’s most and least talented members (and in the process provide the organization with useful data on managers’ ability to spot and champion talent)—it offers independent verification of performance-appraisal data, something we both agree is important.

Sure, there are challenges involved in implementing a forced-ranking system. Some employees and some managers, particularly those with low standards, don’t like it. And it’s inevitable that some mistakes will arise—you’re bound to miss a few late bloomers and overrate a few glib duds.

But combining a forced-ranking system with a conventional performance-appraisal system, with senior managers holding their juniors accountable for excellence in performance management (just as they hold them accountable for excellence in all the other parts of their jobs), will produce an organizational climate in which people know what’s expected of them, are held to high standards, and know exactly how well they’re doing. Sounds like a great place to work!

Best regards,

Dick

Hi Dick—

I have concerns about your forced-ranking comments. One is that the predetermined distribution seems extremely arbitrary. Let’s assume for now that the required ranking can be done without error—that every person can be assigned to the correct performance category. For simplicity, let’s also assume there are one hundred people to be assigned, and that the required distribution is twenty A players, seventy B players, and ten C players. This means that the person ranked #21 will be assigned to the same category as person #90 and, of course, to a category different from #20. I would bet that in most cases the performance and potential of #21 is a good deal more like #20 than #90, and that the same would be true of many of the other B players as well.

So what does it mean if Chris has been designated an A player? It means that Chris ended up somewhere within the percentage that was chosen to distinguish A’s from not-A’s. If some other percentage had been chosen, then maybe Chris would have been a B player. In the absence of some logical rationale for choosing one percentage rather than another, that seems pretty arbitrary.

The same can be said about the bottom of the distribution. In some (though I realize not all) forced-ranking systems, there are severe consequences for being ranked in the bottom category. How can it be known ahead of time that #91 and #92 will deserve those consequences, but #90 will not?

And all of the above assumes that the ranking process is flawless. All of the research points to the
conclusion that it is an imperfect process at best, and
more so as time goes on: If forced ranking succeeds
in improving the workforce’s overall quality, then the
difference between the more and less able employees
will surely shrink, making it even more difficult to
rank employees accurately in the future.

I also wonder about all of the workers who will
see their rankings falling over time, since with every
new A-player hire, somebody must fall out of the A
range. Maybe that’s a healthy situation for the or-
ganization, but I’d like to see the evidence before
accepting that conclusion. Detractors have made a
number of other arguments that I won’t go into here,
and so far there is little empirical evidence to support
their claims—but there is also little to refute them.

Before initiating (or continuing) a forced-ranking
program, we owe it to our employees and to our or-
ganizations to show through a realistic and thorough
assessment that the system’s expected benefits out-
weigh the expected costs.

Best wishes,
Steve

From: Dick Grote
To: Steve Scullen

Greetings, Steve!

I appreciate your concerns about forced ranking.
But as it happens, all of those concerns turn out to be benefits. Certainly, using a predetermined distri-
bution (such as top 20 percent, vital 70 percent, and
bottom 10 percent) is arbitrary—and that’s its great
value. Using fixed and arbitrary percentages forces
managers to make tough decisions about who’s an A
player, who’s not, and why not. Otherwise, as hap-
pens in too many performance-appraisal systems,
everyone gets rated superior, managers never have
to have tough conversations about performance, and
the organization slowly slouches toward mediocrity.
Restricting the number that can fall into the A cate-
gory, and demanding that managers identify a bot-
tom 10 percent who, relative to their peers, are weak
performers, ensures that top talent is recognized and
that those bringing up the rear have no false sense of
security.

You’re undoubtedly right about #21 being much
closer in performance to #20 than she is to #90.

That’s why companies that use the forced-ranking
process tailor the actions they take with individuals
to the individuals themselves, not just to which rank-
ing bucket the person ended up. When I write scripts
for managers to use in letting people know how they
came out in their company’s A-, B-, and C-player
analysis, I develop five scripts, not just three: for the
solid A player, the B+ (your #21 guy and his coun-
terparts), the genuine B, the B- (the ones who barely
avoided falling into the C category), and finally the
true C-level performer.

But it is important to use buckets in making rela-
tive comparisons (e.g., top 20, vital 70, bottom 10;
or quartiling, or some similar scheme). Never ask
managers to precisely rank their people in exact per-
formance order. It’s impossible to distinguish be-
tween #20 and #21, and the totem-pole approach
(who’s #1, who’s #2, and so on down until the last
and worst performer is fingered) generates all the
concerns about accuracy you raise.

Yes, forced ranking is an imperfect process, as is
any process in which fallible human beings must
make tough decisions in an arena where solid,
unarguable, quantitative data don’t exist. The forced-
ranking process requires the exercise of honed, ob-
jective managerial judgment in a situation where in-
formation is always incomplete and the facts are
sometimes contradictory. But managers make deci-
sions based on limited data all the time—which pro-
jects to fund, which to shelve; when to react swiftly
to a competitor’s move, when to let time take its
course. Just because a decision isn’t based on count-
able units doesn’t mean it isn’t objective. Employee
ranking is not the same as solving an algebra prob-
lem—it can’t be reduced to a mathematical formula.

Whether it’s a conventional annual review or the
relative-comparison, forced-ranking process we’ve
been discussing lately, let’s recognize what a per-
formance appraisal is: a formal record of a manager’s
opinion of the quality of an employee’s work. The key
word here is opinion.

Managers are paid to assess and evaluate just
how good a job George has done, both in meeting his
assigned objectives and in comparison with how well
everyone else on the team met theirs. That opinion
needs to be backed up by all the data the manager
can get. It should be supplemented by being sub-
jected to tough-minded senior management review
and to calibration sessions with the manager’s peers.
to assure reliability and validity and provide a level playing field for all.

As I said earlier, every person in every organization wants the answers to two questions: First, what do you expect of me? Second, how am I doing at meeting your expectations? Besides all the business benefits that good performance-management processes provide, they also require leaders at every level to meet their ethical obligation to answer those two questions.

Best regards!
Dick

From: Steve Scullen
To: Dick Grote
Hi Dick—

I fully agree that a totem-pole ranking procedure is unusable for all but the smallest of groups. Moreover, I agree that in most instances, making a meaningful distinction between #20 and #21 is impossible. But doesn’t the forced-ranking procedure (using the distributions we have been talking about) require managers to do just that? They must put #20 into the A bucket and #21 into the B bucket. Similarly, #90 (B bucket) must be separated from #91 (C bucket).

The fact that #21 cannot be in the same bucket as #20 illustrates to me the arbitrariness of a forced distribution. And predetermining that #91 must be in the same bucket as #100 strikes me the same way. Number 91’s performance and potential could be a good deal more like #70’s than like #100’s, and that makes me uncomfortable with a hard and fast rule saying that #91 must be treated like #100 and not like #70.

I agree with you, however, that managers often fail to make distinctions where real differences exist—and this is unfair, especially to the best performers. I have no problem with a thoughtfully designed and well-executed system for identifying high performers, solid performers, etc. But in the absence of a well-articulated rationale for choosing one performance distribution over others, I am troubled by a rigid system that dictates the number of people to be assigned to each category.

I think a better approach would be to create some set of buckets based on meaningful differences between people, where there is something qualitatively different about A players versus people who are not A players. Maybe this could be done on the basis of performance standards. Or maybe the best that could be done is to get management together to decide who should and who should not be included in the top group of performers. Either way, my hope would be that if someone is not categorized as an A player, it is because she has not achieved at the level expected of A players rather than because there can be only so many A’s. Another undesirable possibility is that someone who really hasn’t performed at the highest level might be included in the A group simply because a certain number of names must be placed in that category. At the other extreme, I would hope that if people in the lowest bucket are being terminated, it is because their performance has been unsatisfactory and not simply because they are in the lowest bucket.

To summarize my thoughts on all of the topics we have discussed, I am not unenthusiastic about the theory of using performance appraisal or forced ranking. Clearly, both processes have the potential to improve organizations and the working lives of their employees. But I continue to have reservations based on what actually happens when either of them is put into practice.

High-quality, accurate performance information is crucial to the effective development of both employees and organizations. Unfortunately, a good deal of empirical evidence suggests that most performance-appraisal data is of less value than it could be. You note several things that can be done to improve the reliability, the validity, and therefore the quality of performance appraisal. I believe you’re on the right track, but there is much to be done. I urge decision-makers to demonstrate their commitment to the process by investing their own time, energy, and whatever other resources are needed to do appraisal well. Then they can expect others to do the same, and should hold them accountable for doing so.

Best wishes and many thanks for a provocative discussion,

Steve
ABOUT DICK GROTE

Dick Grote is President of Grote Consulting Corporation in Dallas, Texas. He is a frequent speaker at corporate conferences and meetings. Dick also regularly presents executive overviews of best practices in performance management to senior executives and HR leadership teams.

His management seminars, MAKING PERFORMANCE APPRAISAL WORK and MANAGING THE MARGINAL PERFORMER, help managers use their organization’s performance management processes effectively and solve the everyday problems of poor attendance, substandard performance and bad attitudes that arise in every organization.

He is the author of the books, Discipline Without Punishment and The Complete Guide to Performance Appraisal. Both books were major book club selections and have been translated into Chinese and Arabic. Discipline Without Punishment has become a management classic and has recently been issued in an updated second edition.

Paramount Pictures bought the movie rights to Discipline Without Punishment and produced the video series “Respect and Responsibility” with Dick as host. His highly popular book, The Performance Appraisal Question and Answer Book, was published by the American Management Association in 2002.


For five years, Dick Grote was a regular commentator on National Public Radio’s “Morning Edition” program. For twenty years he was adjunct professor of management at the University of Dallas graduate school. His articles have appeared in the Harvard Business Review and The Wall Street Journal. His biography appears in Who’s Who in America and Wikipedia.

ABOUT GROTE CONSULTING CORPORATION

Grote Consulting Corporation concentrates exclusively in the area of performance management. It is one of America's best-known and most respected specialized management consulting firms. Based in Dallas, Texas, Grote Consulting helps sophisticated organizations, large and small, design and implement best-practice performance management systems. Our clients include some of the most prestigious organizations, public and private, in North America, Western Europe, and Southeast Asia.