

HBR.ORG

Harvard Business Review

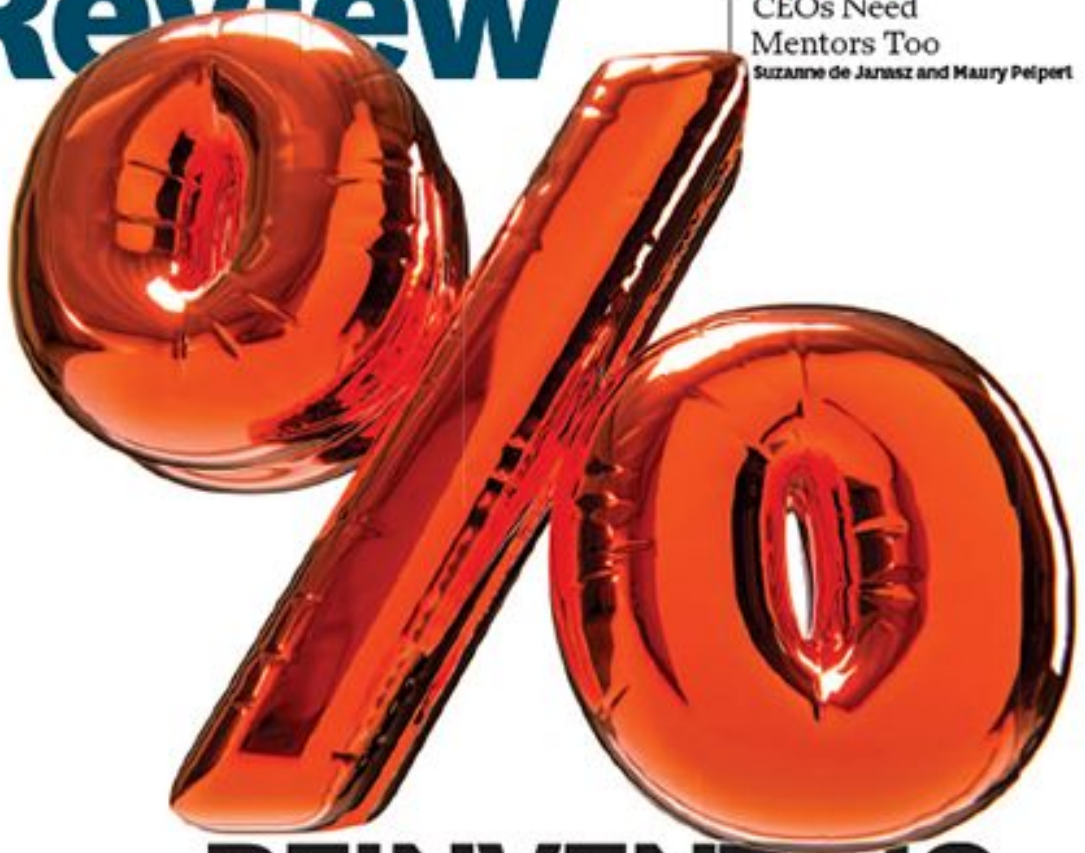


APRIL 2015

63 Spotlight
What Really
Motivates Salespeople

82 Leadership
Family Businesses
and Succession
Claudio Fernández-Arcoz et al.

100 Experience
CEOs Need
Mentors Too
Suzanne de Jarasz and Maury Peiperl



REINVENTING PERFORMANCE RANKINGS

A radical new way to evaluate talent

PAGE 40

Reinventing Performance Management

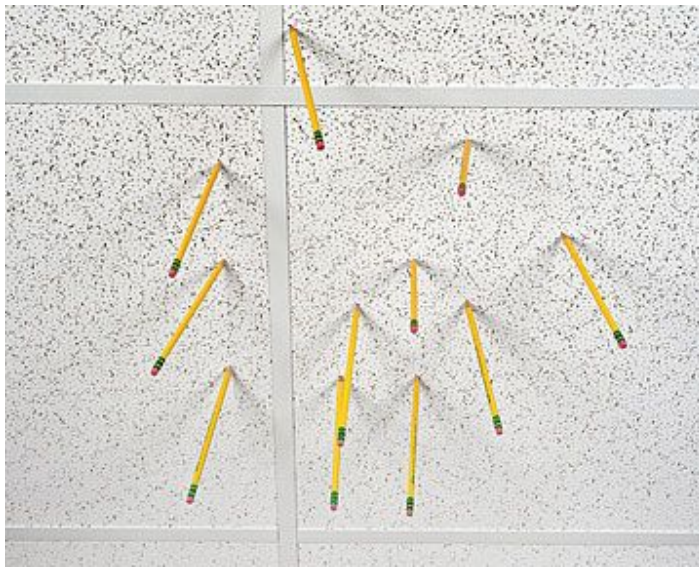
By Marcus Buckingham and Ashley Goodall

Harvard Business Review – April 2015

At Deloitte we're redesigning our performance management system. This may not surprise you. Like many other companies, we realize that our current process for evaluating the work of our people—and then training them, promoting them, and paying them accordingly—is increasingly out of step with our objectives.

In a public [survey](#) Deloitte conducted recently, more than half the executives questioned (58%) believe that their current performance management approach drives neither employee engagement nor high performance. They, and we, are in need of something nimbler, real-time, and more individualized—something squarely focused on fueling performance in the future rather than assessing it in the past.

Further Reading



What might surprise you, however, is what we'll include in Deloitte's new system and what we won't. It will have no cascading objectives, no once-a-year reviews, and no 360-degree-feedback tools. We've arrived at a very different and much simpler design for managing people's performance. Its hallmarks are speed, agility, one-size-fits-one, and constant learning, and it's underpinned by a new way of collecting reliable performance data. This system will make much more sense for our talent-dependent business. But we might never have arrived at its design without drawing on three pieces of evidence: a simple counting of hours, a review of research in the science of ratings, and a carefully controlled study of our own organization.

Counting and the Case for Change

More than likely, the performance management system Deloitte has been using has some characteristics in common with yours. Objectives are set for each of our 65,000-plus people at the beginning of the year; after a project is finished, each person's manager rates him or her on how well those objectives were met. The manager also comments on where the person did or didn't excel. These evaluations are factored into a single year-end rating, arrived at in lengthy "consensus meetings" at which groups of "counselors" discuss hundreds of people in light of their peers.

Internal feedback demonstrates that our people like the predictability of this process and the fact that because each person is assigned a counselor, he or she has a representative at the consensus meetings. The vast majority of our people believe the process is fair. We realize, however, that it's no longer the best design for Deloitte's emerging needs: Once-a-year goals are too "batched" for a real-time world, and conversations about year-end ratings are generally less valuable than conversations conducted in the moment about actual performance.

But the need for change didn't crystallize until we decided to count things. Specifically, we tallied the number of hours the organization was spending on performance management—and found that completing the forms, holding the meetings, and creating the ratings consumed close to *2 million hours a year*. As we studied how those hours were spent, we realized that many of them were eaten up by leaders' discussions behind closed doors about the outcomes of the process. We wondered if we could somehow shift our investment of time from talking to ourselves about ratings to talking to our people about their performance and careers—from a focus on the past to a focus on the future.

We found that creating the ratings consumed close to 2 million hours a year.

The Science of Ratings

Our next discovery was that assessing someone's *skills* produces inconsistent data. Objective as I may try to be in evaluating you on, say, strategic thinking, it turns out that how much strategic thinking *I* do, or how valuable *I* think strategic thinking is, or how tough a rater *I* am significantly affects my assessment of *your* strategic thinking.

How significantly? The most comprehensive research on what ratings actually measure was conducted by Michael Mount, Steven Scullen, and Maynard Goff and published in the *Journal of Applied Psychology* in 2000. Their study—in which 4,492 managers were rated on certain performance dimensions by two bosses, two peers, and two subordinates—revealed that 62% of the variance in the ratings could be accounted for by individual raters' peculiarities of perception. Actual performance accounted for only 21% of the variance. This led the researchers to conclude (in *How People Evaluate Others in Organizations*, edited by Manuel London): "Although it is implicitly assumed that the

ratings measure the performance of the ratee, most of what is being measured by the ratings is the unique rating tendencies of the rater. Thus ratings reveal more about the rater than they do about the ratee.” This gave us pause. We wanted to understand performance at the individual level, and we knew that the person in the best position to judge it was the immediate team leader. But how could we capture a team leader’s view of performance without running afoul of what the researchers termed “idiosyncratic rater effects”?

Putting Ourselves Under the Microscope

We also learned that the defining characteristic of the very best teams at Deloitte is that they are strengths oriented. Their members feel that they are called upon to do their best work every day. This discovery was not based on intuitive judgment or gleaned from anecdotes and hearsay; rather, it was derived from an empirical study of our own high-performing teams.

Our study built on previous research. Starting in the late 1990s, Gallup conducted a multiyear examination of high-performing teams that eventually involved more than 1.4 million employees, 50,000 teams, and 192 organizations. Gallup asked both high- and lower-performing teams questions on numerous subjects, from mission and purpose to pay and career opportunities, and isolated the questions on which the high-performing teams strongly agreed and the rest did not. It found at the beginning of the study that almost all the variation between high- and lower-performing teams was explained by a very small group of items. The most powerful one proved to be “At work, I have the opportunity to do what I do best every day.” Business units whose employees chose “strongly agree” for this item were 44% more likely to earn high customer satisfaction scores, 50% more likely to have low employee turnover, and 38% more likely to be productive.

We set out to see whether those results held at Deloitte. First we identified 60 high-performing teams, which involved 1,287 employees and represented all parts of the organization. For the control group, we chose a representative sample of 1,954 employees. To measure the conditions within a team, we employed a six-item survey. When the results were in and tallied, three items correlated best with high performance for a team: “My coworkers are committed to doing quality work,” “The mission of our company inspires me,” and “I have the chance to use my strengths every day.” Of these, the third was the most powerful across the organization.

All this evidence helped bring into focus the problem we were trying to solve with our new design. We wanted to spend more time helping our people use their strengths—in teams characterized by great clarity of purpose and expectations—and we wanted a quick way to collect reliable and differentiated performance data. With this in mind, we set to work.

Radical Redesign

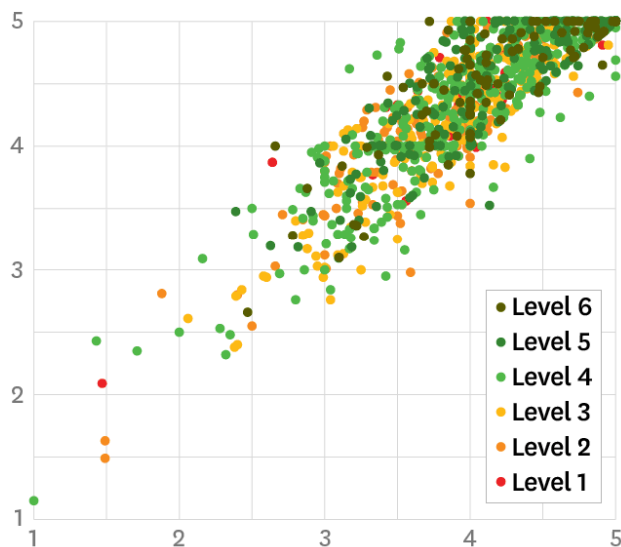
We began by stating as clearly as we could what performance management is actually *for*, at least as far as Deloitte is concerned. We articulated three objectives for our new system. The first was clear: It would allow us to *recognize* performance, particularly through variable compensation. Most current systems do this.

Performance Intelligence

In an early proof of concept of the redesigned system, executives in one large practice area at Deloitte called up data from project managers to consider important talent-related decisions. In the charts below, each dot represents an individual; decision makers could click on a dot to see the person's name and details from his or her "performance snapshots."

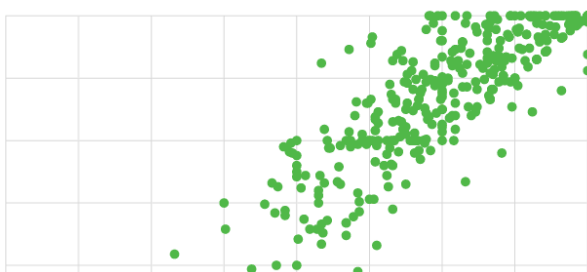
WHAT ARE TEAM LEADERS TELLING US?

First the group looked at the whole story. This view plotted all the members of the practice according to how much their various project managers agreed with two statements: "I would always want this person on my team" (**y axis**) and "I would give this person the highest possible compensation" (**x axis**). The axes are the same for the other three screens.



HOW WOULD THIS DATA HELP DETERMINE PAY?

Next the data was filtered to look only at individuals at a given job level. A fundamental question for performance management systems is whether they can capture enough variation among people to fairly allocate pay. A data distribution like this offers a starting point for broader discussion.



But to recognize each person's performance, we had to be able to *see* it clearly. That became our second objective. Here we faced two issues—the idiosyncratic rater effect and the need to streamline our traditional process of evaluation, project rating, consensus meeting, and final rating. The solution to the former requires a subtle shift in our approach. Rather than asking more people for their opinion of a team member (in a 360-degree or an upward-feedback survey, for example), we found that we will need to ask only the immediate team leader—but, critically, to ask a different kind of question. People may rate other people's skills inconsistently, but they are highly consistent when rating their own feelings and intentions. To see performance at the individual level, then, we will ask team leaders not about the *skills* of each team member but about their *own future actions* with respect to that person.

At the end of every project (or once every quarter for long-term projects) we will ask team leaders to respond to four future-focused statements about each team member. We've refined the wording of these statements through successive tests, and we know that at Deloitte they clearly highlight differences among individuals and reliably measure performance. Here are the four:

1. Given what I know of this person's performance, and if it were my money, I would award this person the highest possible compensation increase and bonus [*measures overall performance and unique value to the organization on a five-point scale from "strongly agree" to "strongly disagree"*].
2. Given what I know of this person's performance, I would always want him or her on my team [*measures ability to work well with others on the same five-point scale*].
3. This person is at risk for low performance [*identifies problems that might harm the customer or the team on a yes-or-no basis*].
4. This person is ready for promotion today [*measures potential on a yes-or-no basis*].

In effect, we are asking our team leaders what they would *do* with each team member rather than what they *think* of that individual. When we aggregate these data points over a year, weighting each according to the duration of a given project, we produce a rich stream of information for leaders' discussions of what they, in turn, will do—whether it's a question of succession planning, development paths, or performance-pattern analysis. Once a quarter the organization's leaders can use the new data to review a targeted subset of employees (those eligible for promotion, for example, or those with critical skills) and can debate what actions Deloitte might take to better develop that particular group. In this aggregation of simple but powerful data points, we see the possibility of shifting our 2-million-hour annual investment from talking about the ratings to talking about our people—from ascertaining the facts of performance to considering what we should do in response to those facts.

We ask leaders what they'd do with their team members, not what they think of them.

In addition to this consistent—and countable—data, when it comes to compensation, we want to factor in some uncountable things, such as the difficulty of project assignments in a given year and contributions to the organization other than formal projects. So the data will serve as the starting point for compensation, not the ending point. The final determination will be reached either by a leader who knows each individual personally or by a group of leaders looking at an entire segment of our practice and at many data points in parallel.

We could call this new evaluation a rating, but it bears no resemblance, in generation or in use, to the ratings of the past. Because it allows us to quickly capture performance at a single moment in time, we call it a *performance snapshot*.

The Third Objective

Two objectives for our new system, then, were clear: We wanted to recognize performance, and we had to be able to see it clearly. But all our research, all our conversations with leaders on the topic of performance management, and all the feedback from our people left us convinced that something was missing. Is performance management at root more about “management” or about “performance”? Put differently, although it may be great to be able to measure and reward the performance you have, wouldn’t it be better still to be able to improve it?

Our third objective therefore became to *fuel* performance. And if the performance snapshot was an organizational tool for measuring it, we needed a tool that team leaders could use to strengthen it.

How Deloitte Built a Radically Simple Performance Measure

One of the most important tools in our redesigned performance management system is the “performance snapshot.” It lets us see performance quickly and reliably across the organization, freeing us to spend more time engaging with our people. Here’s how we created it.

1. The Criteria

We looked for measures that met three criteria. To neutralize the idiosyncratic rater effect, we wanted raters to rate their own actions, rather than the qualities or behaviors of the ratee. To generate the necessary range, the questions had to be phrased in the extreme. And to avoid confusion, each one had to contain a single, easily understood concept. We chose one about pay, one about teamwork, one about poor performance, and one about promotion. Those categories may or may not be right for other organizations, but they work for us.

2. The Rater

We were looking for someone with vivid experience of the individual's performance and whose subjective judgment we felt was important. We agreed that team leaders are closest to the performance of ratees and, by virtue of their roles, must exercise subjective judgment. We could have included functional managers, or even ratees' peers, but we wanted to start with clarity and simplicity.

3. Testing

We then tested that our questions would produce useful data. Validity testing focuses on their difficulty (as revealed by mean responses) and the range of responses (as revealed by standard deviations). We knew that if they consistently yielded a tight cluster of "strongly agree" responses, we wouldn't get the differentiation we were looking for. *Construct* validity and *criterion-related* validity are also important. (That is, the questions should collectively test an underlying theory and make it possible to find correlations with outcomes measured in other ways, such as engagement surveys.)

4. Frequency

At Deloitte we live and work in a project structure, so it makes sense for us to produce a performance snapshot at the end of each project. For longer-term projects we've decided that quarterly is the best frequency. Our goal is to strike the right balance between tying the evaluation as tightly as possible to the experience of the performance and not overburdening our team leaders, lest survey fatigue yield poor data.

5. Transparency

We're experimenting with this now. We want our snapshots to reveal the real-time "truth" of what our team leaders think, yet our experience tells us that if they know that team members will see every data point, they may be tempted to sugarcoat the results to avoid difficult conversations. We know that we'll aggregate an individual's snapshot scores into an annual composite. But what, exactly, should we share at year's end? We want to err on the side of sharing more, not less—to aggregate snapshot scores not only for client work but also for internal projects, along with performance metrics such as hours and sales, in the context of a group of peers—so that we can give our people the richest possible view of where they stand. Time will tell how close to that ideal we can get.

Research into the practices of the best team leaders reveals that they conduct regular check-ins with each team member about near-term work. These brief conversations allow leaders to set expectations for the upcoming week, review priorities, comment on recent work, and provide course correction, coaching, or important new information. The conversations provide clarity regarding what is expected of each team member and why, what great work looks like, and how each can do his or her best work in the upcoming days—in other words, exactly the trinity of purpose, expectations, and strengths that characterizes our best teams.

Our design calls for every team leader to check in with each team member once a week. For us, these check-ins are not *in addition* to the work of a team leader; they *are* the work of a team leader. If a leader checks in less often than once a week, the team member's priorities may become vague and aspirational, and the leader can't be as helpful—and the conversation will shift from coaching for near-term work to giving feedback about past performance. In other words, the content of these conversations will be a direct outcome of their frequency: If you want people to talk about how to do their best work in the near future, they need to talk often. And so far we have found in our testing a direct and measurable correlation between the frequency of these conversations and the engagement of team members. Very frequent check-ins (we might say *radically* frequent check-ins) are a team leader's killer app.

That said, team leaders have many demands on their time. We've learned that the best way to ensure frequency is to have check-ins be initiated by the team member—who more often than not is eager for the guidance and attention they provide—rather than by the team leader.

To support both people in these conversations, our system will allow individual members to understand and explore their strengths using a self-assessment tool and then to present those strengths to their teammates, their team leader, and the rest of the organization. Our reasoning is twofold. First, as we've seen, people's strengths generate their highest performance today and the greatest improvement in their performance tomorrow, and so deserve to be a central focus. Second, if we want to see frequent (weekly!) use of our system, we have to think of it as a consumer technology—that is, designed to be simple, quick, and above all engaging to use. Many of the successful consumer technologies of the past several years (particularly social media) are *sharing* technologies, which suggests that most of us are consistently interested in ourselves—our own insights, achievements, and impact. So we want this new system to provide a place for people to explore and share what is best about themselves.

Transparency

This is where we are today: We've defined three objectives at the root of performance management—to *recognize*, *see*, and *fuel* performance. We have three interlocking rituals to support them—the annual compensation decision, the quarterly or per-project performance snapshot, and the weekly check-in. And we've shifted from a batched focus on the past to a continual focus on the future, through regular evaluations and frequent check-ins. As we've tested each element of this design with ever-larger groups across Deloitte, we've seen that the change can be an evolution over time: Different business units can introduce a strengths orientation first, then more-frequent conversations, then new ways of measuring, and finally new software for monitoring performance. (See the exhibit "Performance Intelligence.")

But one issue has surfaced again and again during this work, and that's the issue of transparency. When an organization knows something about us, and that knowledge is

captured in a number, we often feel entitled to know it—to know where we stand. We suspect that this issue will need its own radical answer.

It's not the number we assign to a person; it's the fact that there's a single number.

In the first version of our design, we kept the results of performance snapshots from the team member. We did this because we knew from the past that when an evaluation is to be shared, the responses skew high—that is, they are sugarcoated. Because we wanted to capture unfiltered assessments, we made the responses private. We worried that otherwise we might end up destroying the very truth we sought to reveal.

But what, in fact, is that truth? What do we see when we try to quantify a person? In the world of sports, we have pages of statistics for each player; in medicine, a three-page report each time we get blood work done; in psychometric evaluations, a battery of tests and percentiles. At work, however, at least when it comes to quantifying performance, we try to express the infinite variety and nuance of a human being in a single number.

Surely, however, a better understanding comes from conversations—with your team leader about how you're doing, or between leaders as they consider your compensation or your career. And these conversations are best served not by a single data point but by many. If we want to do our best to tell you where you stand, we must capture as much of your diversity as we can and then talk about it.

We haven't resolved this issue yet, but here's what we're asking ourselves and testing: What's the most detailed view of you that we can gather and share? How does that data support a conversation about your performance? How can we equip our leaders to have insightful conversations? Our question now is not *What is the simplest view of you?* but *What is the richest?*

Our question now is not *What is the simplest view of you?* but *What is the richest?*

Over the past few years the debate about performance management has been characterized as a debate about ratings—whether or not they are fair, and whether or not they achieve their stated objectives. But perhaps the issue is different: not so much that ratings fail to convey what the organization knows about each person but that as presented, that knowledge is sadly one-dimensional. In the end, it's not the particular number we assign to a person that's the problem; rather, it's the fact that there *is* a single number. Ratings are a distillation of the truth—and up until now, one might argue, a necessary one. Yet we want our organizations to know us, and we want to know ourselves at work, and that can't be compressed into a single number. We now have the technology to go from a small data version of our people to a big data version of them. As we scale up our new approach across Deloitte, that's the problem we want to solve next.

A version of this article appeared in the [April 2015](#) issue (pp.40–50) of *Harvard Business Review*.

Marcus Buckingham provides performance management tools and training to organizations. He is the author of several best-selling books and the forthcoming *StandOut 2.0: Assess Your Strengths, Find Your Edge, Win at Work* (Harvard Business Review Press).

Ashley Goodall is the director of leader development at Deloitte Services LP, based in New York.