

ASSESSING PERFORMANCE

Why Most Performance Evaluations Are Biased, and How to Fix Them

by Lori Mackenzie, JoAnne Wehner and Shelley Correll

JANUARY 11, 2019



DANIEL GRILL/GETTY IMAGES

For many companies, performance review season is kicking off with the new year. Although every organization relies on a different evaluation process, most follow a predictable pattern: First, they invite employees to write about their accomplishments and what they need to improve. Then managers write assessments of their work, offer feedback, and rate their performance on a scale of how well they met expectations.

Underlying this process is the belief that by reflecting on people’s performance and codifying it in an evaluation form, we will be able to assess their merits objectively, give out rewards fairly, and offer useful feedback to help them develop in the next year. But while we may strive to be as meritocratic as possible, [our assessments](#) are imperfect and all too often biased.

As innocuous as the typical form may seem, our research has found that it often allows for our implicit biases to creep in. The problem is the “open box.” Most forms ask managers broad questions about their employees—e.g., “Describe the ways the employee’s performance met your expectations” or “What are their significant accomplishments?”—and offer a blank space or open box that managers can fill with assessments, advice, and criticisms as they see fit.

The ambiguity of these questions is by design: They are general and open-ended precisely because they must apply to everyone in the organization, regardless of level or function. So when the form states “Describe the ways the employee’s performance met your expectations,” managers are expected to remember or figure out on their own what the specific expectations were for that particular employee.

The trouble is, when the context and criteria for making evaluations are ambiguous, bias is more prevalent. As many [studies](#) have shown, without structure, people are more likely to rely on gender, race, and other stereotypes when making decisions – instead of thoughtfully constructing assessments using agreed-upon processes and criteria that are consistently applied across all employees.

And while ambiguity opens the door to bias, our research shows that individuals can take actions to reduce that ambiguity and be more objective when filling in the open box.

Open box = open to bias

Our research team at the Stanford VWare Women’s Leadership Lab conducted in-depth studies of evaluation processes at three companies based in the U.S. We uncovered patterns of ambiguity in how performance reviews are written that can lead to a disadvantage for women.

In analyzing men’s and women’s written performance reviews, we discovered that women were more likely to receive [vague feedback](#) that did not offer specific details of what they had done well and what they could do to advance. Women were more likely to be told, for example, to “do more work in person” with no explanation about the issue to overcome or the goal of the change. Men were more likely to receive longer reviews that focused on their technical skills, compared to shorter reviews for women that were more concerned with their communication skills.

Next we observed some [performance discussions](#) at two companies in what are known as “calibration” or “talent review” meetings. At these sessions, leaders have a fixed period of time—for example, three minutes—to provide rationale for an employee’s rating, and then they discuss and

align their ratings. As with written reviews, these oral presentations had a wide degree of variation in what was covered.

In some, the conversation focused on the employee's accomplishments and strengths. In others, a balanced view was given including opportunities to improve. People varied in what criteria was important or valued, and these patterns of variance often followed [gendered expectations](#). The majority of criticisms of women's personalities were about being too aggressive, where the majority for men's were about being too soft.

The informal format allowed leaders to override one another's presentations with simple phrases like, "The style stuff doesn't matter. He is great, and it is irrelevant." And lack of structure led to very different reviews that tended to advantage men—describing them in ways that align with leadership and providing them the coaching they need to advance, while offering women less praise and less actionable guidance to work with.

This kind of variance in evaluations did not surprise many managers. In one project, only 15% of women and 24% of men managers had confidence in the performance evaluation process, while most viewed it as subjective and highly ambiguous.

“Constrain” the Open Box

At one [site](#), we explored how the team could fix the ambiguity in their performance management. We agreed that an overhaul of their performance review system was not the answer. A new system would not necessarily change how managers wrote evaluations differently for men than for women. Instead, we identified a set of discrete actions that managers could take to make their evaluations fairer and more effective.

Since open boxes in performance review documents were inviting bias, we looked for ways to require specificity in managers' assessments. With the input of the managers, we created a checklist to help them consistently reference specific and predetermined data when filling the open boxes. First, it asked "Did you collect the following evidence/data for this employee over the past 6 months" to ensure that comparable data was collected for all employees. Then the checklist asked managers to use the same criteria for all employees by prompting, "While writing your evaluations did you consider the following (previously agreed criteria)?"

Guided by these questions, managers were able to offer more specific and evidence-based feedback to their employees. At the end of our engagement, 90% of the managers told us that they felt the process helped them be more consistent and fairer. They also felt more confident; one told us, "Before I was lost, I admit. Now I had clear criteria and I was judging everyone the same."

In another site, when managers consistently applied their criteria to employees, there was a reduction in the gender gaps in ratings, eliminating the overrepresentation of men in the top performance category and women in the middle.

You can make your performance reviews fairer and more consistent too, even if your organization does not change the review form. Here are three small, simple, yet impactful, things you can do to “constrain” the open box:

Create a rubric for evaluations. Managers often report that they start writing their evaluations without first reviewing their employees’ original goals or establishing a methodology to ensure the assessments are fair. An effective rubric first defines the criteria against which the employee’s performance will be assessed. Then, it requires taking evidence from the employee’s outcomes to assess whether they did or did not meet expectations.

By first creating a rubric, then filling in the open box with your assessment and feedback, you will be less prone to be influenced by your gut reactions. Research shows that when you first agree to the criteria used in the assessment and then you make the evaluation, you are less likely to rely on stereotypes and your assessments are less biased.

Create better prompts. When writing reviews, managers often vary in what they cover, how much they write, and even how specific or vague the comments are. It might be tempting to think this variation reflects the employee’s actual performance (“He’s great! Of course I have a lot to say about him.”), when in fact it might be implicit bias in action. Better prompts can help you approach each review in a similar manner, ensuring everyone is evaluated and considered in a consistent and equal way. Take the query: “Describe the ways the employee’s performance met your expectations.” To be fairer and more consistent, you might prompt yourself to identify three specific, measurable outcomes for each of your employees.

Run a consistency check. Get in the habit of re-reading all reviews for consistency. Even if you have clarified the criteria and created checklists to guide your assessments, you may still fall into patterns that are more favorable to some employees. By looking for uniformity—or patterns of variation—you may find additional ways to remove bias.

Boosting everyday performance

Constraining the open box can be a tool for formal performance evaluations, but also during everyday interactions. At one mid-sized tech company, we shared this approach for blocking bias with a group of managers. One Dev manager instantly saw how this applied to his weekly one-on-one meetings with his team. He often asked open-ended questions about what support each person needed. Upon reflection, the manager realized that the men tended to ask for his support on technical issues, while the women often asked for guidance on effectively engaging with their teams.

This variance in feedback could lead to different expertise being developed by the men and women on his team, but also to different career trajectories. What he needed was for each employee to be *both* a technical expert and competent team leader. As a result, he said he would no longer use broad, open questions, but specifically prompt his employees about both technical and management issues.

What the Dev manager—and the many people we have engaged—learned is that ambiguity in assessments can lead to bias. They take this insight and find ways to use rubrics and prompts to be consistent and fair. It might be tempting to think we can just trust our instincts. But the challenge is that implicit bias creep in, and it's really difficult to see and therefore stop it in its tracks. These three seemingly simple strategies are powerful precisely because they help us bypass our imperfect and often compromised impressions—and to deliver on our intention to be fair to everyone.

Lori Mackenzie is Executive Director of the Clayman Institute at Stanford University and co-founder of Stanford VMware Women's Leadership Innovation Lab.

JoAnne Wehner, PhD is a sociologist at the VMware Women's Leadership Innovation Lab at Stanford University.

Shelley Correll is professor of sociology and organizational behavior at Stanford University, Director of the Stanford VMware Women's Leadership Innovation Lab, and the Barbara D. Finberg Director of the Clayman Institute for Gender Research.

Copyright 2019 Harvard Business Publishing. All Rights Reserved. Additional restrictions may apply including the use of this content as assigned course material. Please consult your institution's librarian about any restrictions that might apply under the license with your institution. For more information and teaching resources from Harvard Business Publishing including Harvard Business School Cases, eLearning products, and business simulations please visit hbsp.harvard.edu.