

[BIAS INTERRUPTERS] *small steps big change*

Tools for Performance Evaluations

THE CHALLENGE

A study of performance evaluations in tech found that 66% of women's performance reviews contained at least one negative personality criticism ("You come off as abrasive") whereas only 1% of men's reviews did. In our performance evaluation audit at a law firm, we found that people of color and white women were far more likely to have their personality mentioned in their evaluations (including negative personality traits). What's optional for white men (getting along with others), appears to be necessary for white women and people of color. Case in point: 83% of Black men were praised for having a "good attitude" vs. 46% of white men, and 27% of white women were praised for being "friendly and warm" vs. 10% of white men.

Research also shows that white men tend to be judged on their *potential* while "prove-it-again groups" (women, people of color, individuals with disabilities, members of the LGBTQIA+ community, older employees, and first-generation professionals) are judged (or scrutinized) on their *performance*. Small biases can have large effects: According to one study, women received significantly lower "potential" ratings despite higher job performance ratings and this accounted for 30-50% of the gender promotion gap.

New Systems Benefit Everyone

In one Bias Interrupters experiment, described in a [*Harvard Business Review* article](#), we sharply decreased documented forms of racial and gender bias in performance evaluations at a law firm. In year one, our analysis found that only 9.5% of people of color had leadership mentioned in their performance evaluations – over 70 percentage points lower than white women. In year two, we used two simple Bias Interrupters: we redesigned the performance evaluation form to require evidence to justify competency ratings and helped the organization design a one-hour workshop on how to interrupt bias in performance evaluations. The results were dramatic. In year two, 100% of people of color had leadership mentioned – and evaluations that mentioned leadership skills had higher ratings. And constructive feedback increased for employees of *all* demographic groups which highlights an extremely important point: using an evidence-based performance evaluation system benefits everyone.

THE SOLUTION

1. Use Metrics

Data and metrics help you spot problems—and assess the effectiveness of the measures you've taken. Businesses use metrics to help them achieve any strategic goal.

Key metrics:

Do your performance evaluations show consistently higher ratings for majority men than for women, people of color, or other relevant groups?

Do your performance evaluations show consistently higher ratings for in-person workers than remote and hybrid workers?

Do women's ratings fall after they have children? Do employees' ratings fall after they take parental leave or adopt flexible work arrangements?

Do the same performance ratings result in different promotion or compensation rates for different groups?

Keep metrics by: 1) individual supervisor; 2) department; and 3) the organization as a whole.

Collecting Data

Your organization already collects performance evaluations from managers, but you will need to pull this data in a way that allows you to analyze the demographic breakdown of the data.

Keep in mind that your performance evaluation system data may be stored in a different location than demographic information (which typically is collected in HR records created when someone is first hired):

Demographic data of employees:

Race/Ethnicity. This is likely collected from employees when they first apply to work at your organization.

Gender identity. Again, this is likely collected from employees when they first apply to work at your organization.

Performance evaluations data:

Quantitative ratings. All quantitative ratings that are tracked.

Narrative answers: Answers to all open-ended questions.

Interpreting Data

Examining the demographic breakdown of quantitative ratings will help you determine whether there is a pattern of group differences in your performance evaluations.

Looking at the indicators of bias will help you understand precisely how bias is playing out at your organization and provide a path forward.

Quantitative ratings: In an organization where bias is not playing out in performance evaluations, we would expect the average ratings for each demographic group to be roughly equal. If the ratings at your company differ across groups, that could be evidence of bias.

Narrative data: What managers write in evaluations is also important. For example, in a company with prove-it-again bias, we would expect to see white men being described as valuable assets to the company even when they had lower ratings than women or people of color. A company with tightrope bias would see more personality comments for women and people of color, while white men would be more likely to receive praise for leadership skills.

As you interpret your qualitative data, it is important to keep in mind that the objective is not to call out specific managers or even to spot bias in individual evaluations. Instead, we look for *patterns* that typically only become evident when we read a group of evaluations together.

Acting on Data

Depending on the pattern(s) you see in the pre-intervention data, there are two key areas of focus for your structural intervention:

- Revising the written materials connected to performance evaluation
- Providing training to help people combat bias

If your process includes the following, we have Bias Interrupters for them, too:

- Self-evaluations
- Calibration meetings

We have curated Bias Interrupters for each area of focus. Read through the drop-down menu below to determine which Bias Interrupters should be implemented at your organization.

Interpreting Post-Intervention Data

After implementing your chosen interventions, you will want to examine the impact of the changes you have made. There are a few key indicators you should be looking for:

Ratings: Compare your pre-intervention ratings to the post-intervention ratings.

Are you closer to equal ratings for groups? That is a good indicator that your intervention was impactful.

Are you seeing the same issues as before? That is a good indicator that you need to add more bias interrupters.

Are you seeing more, or different issues than before? Interrupting bias is an iterative process – you may need to make several rounds of changes. Consider the menu of options below, and decide whether you want to add in more bias interrupters to different parts of the performance evaluation process.

More level playing field across groups:

Compare your pre-intervention narrative results to the post-intervention narrative results.

Are you closer to equal numbers for groups? That is a good indicator that your intervention was impactful.

Are you seeing the same issues as before? That is a good indicator that you need to add more bias interrupters.

Are you seeing more, or different issues than before? Interrupting bias is an iterative process – you may need to make several rounds of changes. Consider the menu of options below, and decide whether you want to add in more bias interrupters to different parts of the performance evaluations process.

2. Implement Bias Interrupters detailed in the menu below

Designing the Performance Evaluation Form

• Begin with clear and specific performance criteria directly related to job requirements.

Try: "She writes maintainable code, tests her work thoroughly, offers clear and useful suggestions during code reviews, and communicates well with clients to gather requirements," instead of: "She's a great programmer."

• Require evidence from the evaluation period that justifies the rating.

Try: "This year, he did a great job in helping us win X project, writing a clear client proposal that defined a tight scope and communicated our fee structure in a way that was carefully and strategically considered." instead of: "He's great at helping us win projects."

• Consider performance and potential separately for each candidate.

Given the tendency for majority men to be judged on their potential while others are judged on their performance, the two criteria should be evaluated separately.

• Separate personality issues from skill sets for each candidate.

Personal style should be appraised separately from skills because a narrower range of behavior often is accepted from women and people of color. For example, women may be labeled "difficult" for doing things that are accepted in majority men.

• **Level the playing field with respect to self-promotion** by ensuring everyone knows they're expected to do so and that they know how. Distribute our Writing an Effective Self-Evaluation Guide to help. Some groups, notably women, people of Asian descent, and first-generation professionals may be reluctant to self-promote. By equipping all employees with this worksheet, modest and introverted people can benefit as well.

Controlling for Bias in the Process of Filling out Form

• **Don't eliminate your performance appraisal system.**

Eliminating formal performance evaluation systems and replacing them with feedback-on-the-fly creates conditions for bias to flourish.

• **Don't accept global ratings without back-up.**

Require evidence from the evaluation period that justifies the rating. Try: "In March, she gave X presentation in front of Y client on Z project, answered his questions effectively, and was successful in making the sale," instead of: "She's quick on her feet."

In the [performance evaluation experiment at a law firm](#) (described above), we redesigned the form to focus on specific competencies that mattered to the organization and required that evaluators list 3 pieces of evidence to accompany every numerical rating. Doing so minimized the "halo-horns effect:" where white men are artificially advantaged by global ratings because they get halos (one strength is generalized into an overall high rating) whereas other groups get horns (one mistake is generalized into an overall low rating).

• **Combat in-person favoritism.**

With more companies transitioning to hybrid models of work, it is important to ensure that "face-time" in the office doesn't translate to higher ratings on performance evaluations, quicker promotions, and increased compensation. Instead, when assessing employee performance, be sure to use output-based evaluation.

• **Evaluations for remote/hybrid workers should be done through video conference or in-person.**

To prevent any potential misunderstandings, it is important to have context such as facial expressions.

• **Equip yourself and others** involved in the evaluation process by keeping a copy of our [Performance Evaluation Checklist](#) nearby when writing and reviewing performance evaluations.

• **Provide a bounceback.**

Managers whose performance evaluations show persistent bias should receive a bounceback (i.e. someone should talk through the evidence with them).

What's a bounceback? An example: in one organization, when a supervisor's ratings of an underrepresented group deviate dramatically from the mean, the evaluations are returned to the supervisor with the message: either you have an undiagnosed performance problem that requires a Performance Improvement Plan (PIP), or you need to take another look at your evaluations as a group. The organization found that a few people were put on PIPs— but that over time supervisors' ratings of underrepresented groups converged with those of majority men. The organization that used this found that all groups found performance evaluations equally fair.

The Center for WorkLife Law conducted an experiment with Dr. Monica Biernat at the University of Kansas examining the effects of reading our Identifying Bias in Performance Evaluations Guide. Participants completed reviews for hypothetical employees. Half of the participants were randomly assigned to a group that read the Bias Guide and listened to a brief audio recording summarizing the main messages; the other half received no further instructions.

Our findings indicate that reading the toolkit leads participants to give higher ratings, monetary bonuses, and promotion recommendations for both women and Black workers.

Before the next round of performance evaluations, have everyone on your team watch this short [2 minute video](#) and read the [Identifying Bias in Performance Evaluations Guide](#).

Calibration Meetings



In many organizations, managers meet to produce a target distribution of ratings or cross-calibrate rankings. Adding structure to these meetings can help you avoid common pitfalls.

- 1) Have managers read our [Identifying Bias in Performance Evaluations Guide](#) before they meet.
- 2) **Pre-commit.** Require all managers to fill out and submit their evaluations before they walk into the room. Registering responses in this way ensures that all managers feel empowered to speak up, and opinions won't be swayed based on the evaluation of whoever speaks first.
- 3) **Use a consistent rubric.** Establishing key competency criteria will ensure that you evaluate each employee on the same job-relevant dimensions.
- 4) **Stick to it.** If the conversation strays away from the established competency criteria, steer everyone back to what is relevant. For example, if an employee's personality is brought up, you can say "is this relevant to the rubric?"
- 5) **Have Bias Interrupters play an active role.** Have a trained Bias Interrupter in the room who can take responsibility for ensuring that the conversation sticks to the established criteria. To better understand the research and rationale behind the suggested bias interrupters, read and distribute our [Identifying Bias in Performance Evaluations Guide](#) which summarizes numerous studies.

Writing an Effective Self-Evaluation

Some people feel more comfortable with self-promotion than others. This partly depends on how you were raised: some people were taught to be forthcoming about their accomplishments. Others grew up with the "modesty mandate"—to be self-effacing and underplay their accomplishments.

Individuals can

- **Self-promote effectively** by using our [Writing an Effective Self Evaluation Guide](#), and handing it out to your reports (if you have them).

Managers and organizations can

- **Level the playing field** with respect to self-promotion by ensuring everyone knows they're expected to do so and that they know how. Distribute our Writing an Effective Self-Evaluation Guide to help.
- **Offer alternatives to self-promotion.** Encourage or require managers to set up more formal systems for sharing successes, such as a monthly email that lists employees' accomplishments.

BIAS INTERRUPTERS *small steps big change*

BIAS INTERRUPTERS FOR PERFORMANCE EVALUATIONS *Identifying Bias in Performance Evaluations Guide*

The five patterns below describe *tendencies not absolutes*. Here's what to watch out for:

Prove-It-Again! ("PIA") — Groups stereotyped as less competent often have to prove themselves over and over. "PIA groups" include women, people of color, individuals with disabilities, older employees, members of the LGBTQIA+ community, and first-generation professionals.

1. **"He'll crush it"; "They're not ready."** PIA groups are judged on performance; others on potential.
2. **"He's skilled; she's lucky."** PIA groups' successes attributed to luck, majority men's successes attributed to skill.
3. **"It could happen to anyone"; "She blew it."** PIA groups' mistakes tend to be noticed more and remembered longer, whereas majority men's mistakes tend to be written off.
4. **PIA groups get horns; others a halo.** Horns=one weakness generalized into an overall negative rating. Halo=one strength generalized into a global positive rating. In addition, mistakes by one PIA group member may reinforce negative group stereotypes.
5. **"We applied the rule—until we didn't."** Objective requirements applied rigorously to PIA groups—but applied leniently or waived for majority men.
6. **Little white lies.** Women, people of color and people with disabilities tend to receive less specific and less honest feedback — meaning they aren't given the opportunity to improve.
7. **Do only the superstars survive?** Superstars may escape PIA problems that affect others.

Tightrope — A narrower range of workplace behavior often is accepted from women, people of color, and members of the LGBTQIA+ community ("TR groups"). First-generation professionals and modest or introverted men can face Tightrope problems, too.

1. **Leader or worker bee?** TR groups face pressure to be "worker bees" who work hard and are undemanding...but if they comply, they lack "leadership potential."
2. **Modest, helpful, nice; dutiful daughter, office mom?** Prescriptive stereotypes create pressures on women to be modest, mild-mannered team players—so "ambitious" is not a compliment for women and niceness may be optional for men but required of women.
3. **Direct and assertive—or angry and abrasive?** Direct, competitive, and assertive in majority men may be seen as inappropriate in TR groups — "tactless," "selfish," "difficult," "abrasive." Anger that's accepted from majority men may be seen as inappropriate or even threatening in TR groups.
4. **Office housework vs glamour work.** TR groups report less access to career-enhancing opportunities and more "office housework"—planning parties & cleaning up; taking notes & arranging meeting times; mentoring & being the peacemaker.
5. **"She's a prima donna"; "He knows his own worth."** The kind of self-promotion that works for majority men may be seen as off-putting in TR groups. Modest men may encounter bias about how "real men" should behave. Strong modesty norms can make first-generation professionals, people of Asian descent, and women uncomfortable with self-promotion.
6. **LGBTQIA+ employees** may be stereotyped as "too feminine," "too masculine," or just "too gay." These kinds of judgement signal illegal discrimination under federal and state law.

The Parental Wall can affect parents of all genders — as well as employees without children.

1. **"He has a family to support."** Fathers face expectations that they will not—or should not—take time off for caregiving. They may be seen as deserving more pay or promotion because of their presumed family role.
2. **"Her priorities lie elsewhere."** Mothers are stereotyped as less competent and committed, are held to higher performance and punctuality standards—and are half as likely to be promoted as identical candidates without children.
3. **"I worry about her children."** Mothers who work long hours tend to be disliked and held to higher performance standards.
4. **"It's not a good time for her."** Opportunities or promotions may be withheld on the assumption that mothers will not—or should not—want them.
5. **"No life."** Employees without children may face the assumption that they can always pick up the slack because they have "no life."

Tug of War — Sometimes bias creates conflict within historically excluded groups.

1. **Tokenism.** If there is only one "token" woman or person of color on a team, they might be treated like they are only in the room to provide diversity, not expertise. Then, when it comes times for performance evaluations, they will get dinged for not having completed as many impressive assignments.
2. **Strategic distancing and the loyalty tax.** People from historically excluded groups may feel they need to distance themselves from others of their group, or align with the majority against their own group, in order to get ahead. In performance evaluations, this might mean giving a lower rating to your own group members to avoid being accused of favoritism.
3. **Passthroughs. PIA:** People from historically excluded groups may hold members of their own groups to higher standards because "That's what it takes to succeed here." **Tightrope:** Women or LGBTQIA+ employees may fault each other for being too masculine—or too feminine. People of color may fault each other for being "too white"—or not "white" enough. **Parental wall:** Parents may fault each other for handling parenthood the wrong way—for taking too much time off or too little.

Racial Stereotypes — People of Asian descent are often stereotyped as passive and lacking in social skills; Black people as angry or too aggressive; Latino/a people as hotheaded or emotional.

Seven Powerful Bias Interrupters

1. Give 3 pieces of evidence (from the evaluation period) to explain and back up your rating.
2. Make sure to give everyone—or no one—the benefit of the doubt.
3. If you waive objective rules, do so consistently.
4. Don't insist on likeability, modesty, or deference from some but not others.
5. Don't make assumptions about what mothers—or other caregivers—want or are able to do.
6. If you comment on "culture fit," "executive presence," or other vague concepts, start with a clear definition and keep track to ensure such concepts are applied consistently.
7. Give honest feedback to everyone who is evaluated—otherwise some groups won't get notice of problems in time to correct them.

BIAS INTERRUPTERS

*small steps
big change*

Performance Evaluation Checklist

- 1. Before you start** make sure your evaluation form and process meets the following criteria.
 - △ As a reviewer make sure you have clear criteria of what you're evaluating given the employee's role. Share these criteria with the employee being evaluated ahead of time.
 - △ Designate separate boxes for feedback about performance and potential.
 - △ Designate separate boxes for any personality-related feedback, it should not be clumped in with skillset related remarks.
 - △ If there is a self-review component, be sure to distribute the **Writing an Effective Self-Evaluation Guide** to the employee ahead of time.

- 2. While you review** keep in mind the following:

A recent study of performance evaluations in tech found that 66% of women's performance reviews contained negative personality criticism ("You come off as abrasive") whereas only 1% of men's reviews did.¹ That said, a comment doesn't need to be negative in order to be harmful. For example, when a reviewer says something like "he's friendly and gets along well with everyone," they're leaving out the skills that leave that impression. It's more effective and accurate to instead say, "he has strong interpersonal skills, best showcased by his ability to lead projects across multiple departments." Instead of focusing on someone's personal vibe, zeroing in on their actions will help clarify your feedback and in turn help the employee better address their strengths and weakness.

Flag Words and Phrases in Performance Evaluations:

If one of the terms below is used, reviewers must attach a concrete example/reason for that evaluation. Providing evidence for any assessment should be expected.

- Abrasive, difficult, rude, aggressive
- Personality
- Loud, outspoken
- Quiet, shy
- Tone issues
- Bubbly, perky
- Not a team player
- Not flexible, unreliable
- Unprofessional or inappropriate attire
- Potential, not ready, too green (for promotion, task, etc.)
- Mother, father, parent, their child, caregiver

Some of these phrases may be appropriate in a performance evaluation, but they *must* be contextualized and supported with skills and role criteria-related evidence.

Carve out time for bouncebacks:

If someone you supervise turns in a performance evaluation that doesn't follow the guidelines you've outlined, then consider a bounceback. Take the time to talk through why your department puts a value on the given criteria and create a plan to remedy their evaluations going forward, including editing the one that sparked your concern.

¹ Snyder, K. (2014, August 26). The abrasiveness trap: High-achieving men and women are described differently in reviews. *Fortune*. Retrieved from <http://fortune.com/2014/08/26/performance-review-gender-bias/>

BIAS INTERRUPTERS *small steps big change*

BIAS INTERRUPTERS FOR PERFORMANCE EVALUATIONS

Writing an Effective Self-Evaluation Guide

Some people feel more comfortable with self-promotion than others. Partly it depends on how you were raised: some people were taught to be forthcoming about their accomplishments. Others grew up with the “modesty mandate”—to be self-effacing and underplay their accomplishments.

Regardless of how you were raised, it’s important to learn to engage in deft self-promotion. Research has shown that professionals who self-promote generally are more successful than those who don’t. When asked to provide a self-evaluation as part of your organization’s performance evaluation process, the first step is to understand what’s expected.

Here are a few simple steps to help. They’ll prove useful for all groups, helping you to tone it down a bit if you tend towards bluster or to step it up if you were raised with the modesty mandate.

1. Don’t hold yourself to higher standards: “Am I ready?”

Research shows that women often hold themselves to much stricter standards than men, and often underestimate their own contributions. Women also are less likely to question negative feedback. The same may be true of other groups. Make sure you have a sufficient network at your organization so you know how other people are judging themselves. It’s not fair to yourself—or your organization—to judge yourself too harshly. You are not perfect—but no one else is, either.

2. Identify what objective metrics are important in your context.

If you have a sales goal, it’s easy to identify what objective metrics matter at work. Often metrics are subtler, but most workplaces have objective metrics that matter. Do you know what they are? If not, ask someone you trust. (If you don’t have anyone to ask, ask someone in Human Resources—and work on building your network within your organization.)

3. Show that you’ve met and exceeded the objective metrics that matter.

It may be time-consuming to identify and gather the relevant evidence to show you’ve met the objective metrics that matter. Spend the time. Start collecting this evidence early in the relevant review period. Keep a folder with relevant accomplishments and create documents to track progress.

4. Keep track of compliments.

“Good job”—that’s not the kind of compliment we’re talking about. But if someone told you something like, “That’s the best memo I’ve seen in years from a first-year associate,” you need to find some way of getting that into your self-evaluation. Try this: “I know I still need to develop my skills in X area, but when Y said Z it gave me the confidence to believe my efforts will succeed.”

5. Lose the doubt raisers.

Everyone has doubts. But if you don't show confidence in yourself, why should others show confidence in you? Reread your self-evaluation and take out phrases such as "I just wanted to tell you," "I think maybe," "I wonder if..." Display quiet confidence that you can succeed with hard work. (If that's not true, you can leave later—but even if you leave you want to leave a good impression.)

6. No need to brag; just state the facts.

It's generally better to avoid characterizing yourself and just provide the facts. "I was pleased to be asked to work with team X so early in my career" versus "I am just that brilliant that I am the most junior person ever invited to work on team X." Let others draw the conclusions. Research shows that people who engage in blatant self-promotion tend to put others off.

7. Be comprehensive.

Take the time to include all the facts that make you look good.

8. Be honest and specific about your drawbacks—and how you intend to address them.

Someone who believes they have no room for improvement is difficult to work with. Nobody's perfect; someone who thinks so is deluding themselves. *"I believe that the next step for me is to develop my own accounts, and look forward to working on that next year." "I had a few instances this year where my proofing was not as accurate as it should be, and I have taken steps to ensure that never happens again."* Ideally show that you're making progress in areas of improvement by describing corrective steps you've taken and providing evidence of improvement (if possible). For example:

"I had a few instances this year where I missed meetings due to calendaring errors and received feedback I needed to improve in this area. I implemented a new calendaring system to ensure more reliable schedule-keeping and have not missed any meetings since. I will continue to ensure my attendance meets expectations."

9. What if you made a big mistake?

Be straightforward, acknowledge the mistake, and explain the steps you've taken to make sure it never happens again. But first be sure it *is* a big mistake. Apologizing over and over again for the kinds of small mistakes everyone makes at one point or another shows as poor judgment, as does failing to sufficiently apologize when one does make a Big One.